Author:
Jones, David

Title:
Improving Engineering Information Access and Knowledge Discovery through Model-Based Information Navigation

Improving Engineering Information Access and Knowledge Discovery through Model-Based Information Navigation

David Edward Jones

A thesis submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Mechanical Engineering
July 2018

52,426 words

Abstract

An organisation's data, information, and knowledge is widely considered to be one of its greatest assets. As such, the capture, storage and dissemination of this asset is the focus of both academic and organisational efforts. This is true at the Airbus Group, the industrial partner of this thesis. Their Knowledge Management team invests in state-of-the-art tools and techniques, and actively participates in research in a bid to maximise their organisation's reuse of knowledge and ultimately their competitiveness.

A successful knowledge management strategy creates a knowledgeable and wise workforce that ultimately benefits both the individual and the organisation. The dissemination of information and knowledge such that it is easily and readily accessible is one key aspect within such a strategy. Search engines are a typical means for information and knowledge dissemination yet, unlike the Internet, search within organisations (intranet or enterprise search) is frequently found lacking. This thesis contributes to this area of knowledge management.

Research in the field of enterprise search has been shown to improve search through the application of context to expand search queries. The novel approach taken in this thesis takes this context and applies it visually, moving the search for information away from a text-based user interface towards a user interface that reflects the function and form of the product. The approach: model-based information navigation, is based on the premise that leveraging the visual and functional nature of engineers through a model-based user interface can improve information access and knowledge discovery.

From the perspectives of information visualisation, engineering information management, product life-cycle management, and building information modelling, this thesis contributes through:

- The development of techniques that enable documents to be indexed against the product structure
- The development of techniques for navigation within engineering three-dimensional virtual environments
- The design of a range visual information object for the display of information within engineering three-dimensional virtual environments
- The determination of the affordance of a model-based approach to information navigation

This thesis presents the development of a framework for model-based information navigation: a novel approach to finding information that places a three-dimensional representation of the product at the heart of searching document collections.

Acknowledgements

First and foremost, I am grateful for the supervision and unwavering support provided by Professor Ben Hicks throughout the PhD. Thank you Ben.

Thanks also to Professor Chris McMahon for his support in the role of secondary supervisor during the first couple of years and to Dr. Chris Snider for taking over and helping us over the finishing line.

This thesis was part funded by the Airbus Group through their Knowledge Management team. I would like to thank Airbus and the staff who supported this work: the Knowledge Management team and the Wing In-Service team, and amongst these people, I'd like to single out Nicolas Chanchevrier and Yifan Xie. Both of whom provided support, advice, and access to people, data, and the real-world use cases that underpin this contribution.

Funding was also provided through an EPSRC CASE AWARD and the LOCM project (EP/K014196/2). Many thanks.

The studies presented in this thesis relied on surrogate information and participants provided by the Formula Student Racing teams at the University of West England, the University of Bristol, the University of Bath and Imperial College London. Thank you to each of the teams for their participation and the academic leads, Jason Matthews, Jason Yon, and Jeff Barrie for their support.

In my time at the University of Bristol, I've seen the Design and Manufacturing Futures Lab grow into a strong and dynamic research group full of great people. I would like to thank each and every one of them, both past and present and in particular, Dr. James Gopsill for his enthusiasm, support, and advice.

Thank you to my partner, Lowri, for being by my side, for the perspective, for all the support, and for the proofreading! Finally, to my family. My mum, dad, sister Patsy, and brother John. Completing this thesis has involved a lot hard work and perseverance, and I learned that from you guys. Thank you.

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... [redacted] .... DATE: 2019-01-28 ....

Contents

List of Publications

[-] Jones, D. E., Snider, C., Yon, J., Gopsill, J., Xie, Y., Chanchevrier, N., & Hicks, B. *THE DESIGN OF VISUAL INFORMATION OBJECTS IN THREE-DIMENSIONAL VIRTUAL ENVIRONMENTS FOR ENGINEERING INFORMATION NAVIGATION. In DS92: Proceedings of the DESIGN 2018 15th International Design Conference (pp. 1583-1594).* (2018).

[-] Jones, D., Matthews, J., Xie, Y., Gospill, J., Dotter, M., Chanchevrier, N., & Hicks, B. *Improving engineering information retrieval by combining TD-IDF and product structure classification. In DS 87-6 Proceedings of the 21st International Conference on Engineering Design (ICED 17) (Vol. 6, pp. 6-41).* (2017, August).

[-] Jones, D. E., Xie, Y., McMahon, C., Dotter, M., Chanchevrier, N., & Hicks, B. *Improving Enterprise Wide Search in Large Engineering Multinationals: A Linguistic Comparison of the Structures of Internet-Search and Enterprise-Search Queries. In IFIP International Conference on Product Lifecycle Management (pp. 216-226). Springer, Cham.* (2015, October).

[-] Jones, D. E., Chanchevrier, N., McMahon, C., & Hicks, B. *A strategy for artefact-based information navigation in large engineering organisations. In DS 80-10 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 10: Design Information and Knowledge Management Milan, Italy, 27-30.07. 15.* (2015).

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The data, information and knowledge generated during an organisation's day-to-day activities are widely considered to be one of an its greatest assets [4]. The effective use and re-use of this data, information and knowledge is considered best practice in business management and a cornerstone to competitiveness. This is especially true of the industrial partner in this thesis, the Airbus Group, who invest in the management of data, information and knowledge through their Knowledge Management team.

Data, information, knowledge, and wisdom can be thought of as a layered triangle, see Figure 1.1, the DIKW hierarchy [5]. Data is raw unstructured and unprocessed (with no context/meaning), e.g. numerical production scores. Information is data formatted for a specific purpose, e.g. graphs, CAD drawings, etc. Knowledge is learned information such that it can be applied, e.g. the reading of assembly instructions and the use of the information contained to physically assemble an artefact. Wisdom is then the accumulation of knowledge such that one can apply knowledge from one domain to another, e.g. the ability to assemble an artefact based on the accumulation of knowledge from the use of assembly instructions to assemble a number of other artefacts. It can then be surmised that the more knowledgeable or wise a workforce is the more able they are to perform, benefiting the whole organisation.

The fields of knowledge management and engineering information management systems are specific IT systems aimed at facilitating the capture, storage and dissemination of data, information and knowledge. They exist with the purpose of facilitating the progression up the DIKW hierarchy [5] and towards a knowledgeable workforce. Search

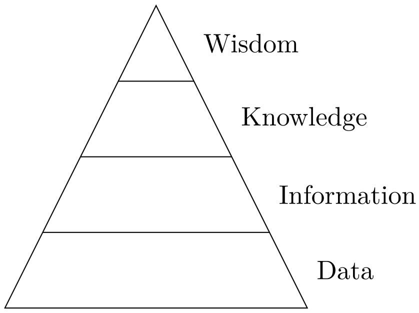


Figure 1.1: The DIKW hierarchy

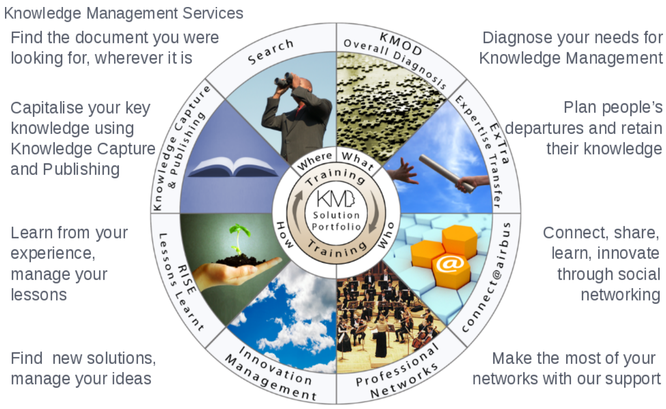


Figure 1.2: The Airbus Knowledge Management Wheel

engines, content management systems and Wikis being examples of general knowledge management systems. Product life-cycle/data management and building information modelling systems being examples of engineering specific systems. These types of system provide the means to structure and store data, information, and knowledge such that they can be widely disseminated across organisations and life-cycles.

The Airbus Knowledge Management team are responsible for the facilitation of knowledge management at the Airbus Group. They achieve this through a range of bespoke knowledge management systems and tailor-made services all aimed at specific areas of the Airbus organisation and their knowledge management needs. Figure 1.2 shows the Airbus Knowledge Management Wheel, the full portfolio of the knowledge management systems and services provided by the Knowledge Management team. In a bid to capitalise on the latest technologies and science, the Knowledge Management team fund research projects in the area of knowledge management and it is within this remit that this thesis stands.

Tasked with aligning with an Airbus function and improving information access and knowledge discovery, this thesis contributes to the field of information and knowledge dis-

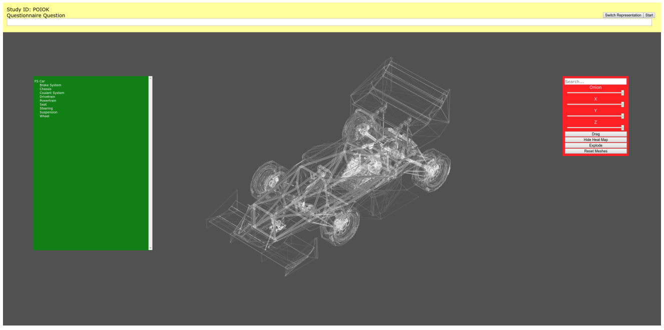


Figure 1.3: Model-based information navigation

semination within large engineering organisations. This contribution is achieved through the design and development of a novel approach to engineering information dissemination: model-based information navigation, see Figure 1.3. A system that allows the navigation of document collections via three-dimensional representation of the product (i.e. a CAD model).

1.2 Thesis Research Context

While a general investment in knowledge management activities and technologies can be considered good business practice, there are a number of challenges specific to Airbus and the aerospace sector that are exemplified in some Airbus functions over others. The Wing In-Service function based at Filton, Bristol, UK, is one such example. Airbus supplements the design, manufacture and sales of aircraft with a service business model[1]. The In-Service functions respond to maintenance and repair requests from customer airlines, with the Wing In-Service focusing on any problems relating to the wings of a specific family of aircraft. The A320 and A380 for example.

To illustrate the Wing In-Service function with an example, an airline will contact one of a number of contact centres with a query such as corrosion on a portion of the wing. The contact centre will verify that the damage is of an extent that falls outside predetermined safety levels and therefore requires a repair designed and tested against the high standards of airworthiness checks. If a repair design is needed the query is passed on to a team of engineers who assess the damage, design a repair, test the repair, and reply to the customer with repair instructions. All this takes time, something that is costly in the aerospace industry. Aircraft that are not flying cost airlines money, and aircraft that are stuck at airport (because it is unclear whether they are safe to fly)

[1] `services.airbus.com/`. Last visited 2018-07-21

incur expensive fines. Speedy response times are then critical for customer satisfaction and, by extension, Airbus's competitiveness. In the case of the In-Service teams, the re-use of information can dramatically reduce the response time, as past repairs have been through safety checks and can be re-used. As such, the In-Service team have a time-sensitive, customer-facing role.

In addition to this and in the wider context of the continuous design and improvement of products, aircraft are long-life, high-value, low-volume, and safety-critical. The Airbus A380 is expected to be in service for over 25 years[2], in 2017 the list price was approximately 436.9 Million USD[3], as of October 2017, the Airbus Group had sold 317[4], and the process of certification by the European Aviation Safety Agency (EASA) and the United States of America's Federal Aviation Administration (FAA) took over five years and encompassed 2600 flight hours with five test aircraft[5].

The re-use of information surrounding aircraft is then vital for both organisational competitiveness and safety, yet there are relatively few opportunities to learn from past and existing products. This is in essence the starting point of this thesis: an alignment with the Wing In-Service team and a project to improve engineering information access, resulting in an improvement day-to-day operation, competitiveness and safety.

There are a range of engineering information management systems and techniques that aim to improve the access and to information and support the engineer in day-to-day operations. These include product data management, product life-cycle management, and building information modelling systems which sit alongside more traditional information technology systems such as content management systems, search engines, and databases. Research in improving search has tended to focus on supplementing search with context, the understanding that when one searches for the term '*cat*', they mean a catalytic converter rather than the animal. Examples of this include the use of structures such as ontologies [6] and taxonomies [7], where the relationships between terms are used to expand and refine search results. Using the information that a catalytic converter is part of an exhaust system and using that to prioritise some results over others for example. Airbus recently implemented the context aware search engine Daedalus within the In-Service function [8].

[2] `http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/test-programme-and-certification/`. Last visited 10th of November 2017.

[3] `http://www.airbus.com/content/dam/corporate-topics/publications/backgrounders/Backgrounder-Airbus-Commercial-Aircraf-price-list-EN.pdf`. Last visited 10th of November 2017.

[4] `http://www.aircraft.airbus.com/aircraftfamilies/passengeraircraft/a380family/`. Last visited 10th of November 2017.

[5] `http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/test-programme-and-certification/`. Last visited 10th of November 2017.

Enterprise or intranet search in particular is a field found lacking. In a study, Stocker et al, [9] introduced Microsoft SharePoint 2013 into two research and development departments, both in engineering organisations (automotive and rail). The findings found problems in three areas: users, documents, and the search engine itself. Users were said to have had problems formulating queries and tagging documents with meta-data when uploading them to the system. Documents were said to be to inconsistent in content and structure, SharePoint itself was said to rank search results misleadingly, with an example of this was ranking by popularity and quarterly finance reports. Users search for the latest/current report however, the previous quarters report will always rank higher by popularity based on the length of time that the document has existed.

Mukherjee et al. [10] explore the possible reasons behind the difficulty in delivering enterprise search when compared to internet search. Many enterprise search systems base themselves on Internet technologies and processes, forgetting the fact that when someone creates a website, they do so wanting search engines to index their content and make it available to all to find. HTML, the language of the Internet, is structured such that website content is easily machine-readable and easy for search engines providers to catalogue. Compare this to a finance team's quarterly reports generated in Microsoft Excel, a two-dimensional technical drawing created in AutoCAD, or some complex computational fluid dynamics analysis of the structure of the wing which are designed with other, non-search related functionality in mind. So these file formats exist for a purpose in their own rights, which do not in any way include being found on the Internet, and yet it is Internet technology that organisations rely on.

While some of these systems have been shown to make improvements over that which came before, Daedalus being a prime example, the problem is still yet to be solved. A study that observed the use of a content management system within a rail and an automotive organisation [9], found the system fell short of expectations. The authors placed the blame on both the users and the system itself, saying that users failed to use the system properly by not adding meta-data, or not knowing the correct search queries to use for example. The system was also quoted as not being able to rank search results appropriately.

Internet technology exists to make any type of topics available to users with any type of interest: it is a one-size-fits-all approach. Systems such as context aware search leverage the context and structures of organisations and improve search. These systems demonstrate that there are improvements in search that can be made if one is to develop systems from the stand point of domain specificity and be open to abandoning some of the generality that the Internet is founded on.

There is therefore scope for further improvement in the area of engineering information management both in terms of search engines and in terms of the user participation. This is the approach taken in this thesis, that a better understanding of the engineer, engineering activities, the product and/or the organisation should logically produce improvements in the way in which engineers find and access information.

1.3 Aim and Research Questions

Engineers think visually and functionally and yet the traditional means of accessing documents (and the information/knowledge contained within), are text-based search engines, speaking to colleagues and traversing shared folder structures. Compare this to the tools that typically support engineers in their work, CAD for example. In recent years, CAD has taken advantage of improvements in computer hardware and software to lift from two-dimensional line drawings to full three-dimensional renderings, arguably moving towards a medium that is far more visual and functional in nature and more akin to the engineer. Within the context of knowledge management, this then begs the question: can a system for the dissemination of information based on the visual and functional form of the product, improve the way engineers access information and discover new knowledge? This leads on to the aim for this thesis: *The development of a framework for model-based information navigation and knowledge discovery in large engineering organisations.*

The concept of accessing data or information via a three-dimensional product representation is not new. Model-based definition [11], annotated/extended CAD [12][13][14] and Building Information Modelling (BIM) [15] are all pre-existing examples. Model-based definition and annotated/extended CAD augment the CAD model such that information such as design rationale is captured within the CAD model itself. Building information modelling makes building related data and information, thermal performance and asset lists for example, available within a three-dimensional representation of the building itself. These examples focus on the storing of information or data directly within the product structure. Going back to the first paragraph of this chapter, so much of an organisation's information and knowledge are captured within documents and databases and, within the case of Airbus, made accessible via a text-based search engine. There is then a gap in the knowledge management systems available to engineering organisations, and there is also a gap in the science in how one would design and deliver such a knowledge management system, the appropriate design decisions and the affordance of such an approach to information navigation.

The combination of the scoping study, literature review and prototype development, lead to a set of research questions being developed. The purpose of these questions being either the identification of the appropriate techniques for developing a model-based approach to information navigation, or to determine the affordance of such an approach.

The research questions are then as follows:

RQ 1 What are the most appropriate techniques for a model-based approach to document indexing?

RQ 2 What are the most appropriate techniques for navigation information within a model-based virtual environment?

RQ 3 What are the most appropriate techniques for displaying information within the model-based virtual environment?

RQ 4 How does model-based information navigation improve engineering information access and knowledge discovery?

The first research question relates to how one indexes documents against the product structure such that documents appear in the appropriate place. The second research question relates to the techniques required to navigate a three-dimensional virtual environment representation of the product/model and find information. The third research question relates to how information should be displayed such that it is able to be found within the three-dimensional model-based virtual environment. These first three research questions form the framework outlining an approach to model-based information navigation. The fourth and final research question then assesses the whether such a system and approach to information navigation improves information access and navigation.

The thesis structure diagram in the next section, Figure 1.5, shows how three first three of these research questions corresponds to a thesis chapter, and they form the framework required to develop a model-based approach to information navigation. The fourth research question is divided up between two chapters, the outlining a study design and the second presenting the study results.

1.4 Thesis Structure

Figure 1.4 shows the approach applied in this thesis. Using the Design Research Methodology [1], this thesis is divided into four stages: Research Clarification, Descriptive Study 1, Prescriptive Study, and Descriptive Study 2. Research Clarification uses a combination of literature review and stakeholder meetings to generate the research aim. The Descriptive Study 1 stage uses a further literature review, a scoping study and an exploration of technology to generate four research questions. The Prescriptive Study uses literature review and studies to answer the first three research questions which in turn produce a framework for model-based information navigation. Descriptive Study 2 then evaluates the model-based approach to information navigation through a further study developed around a system constructed using the framework. This methodology is discussed in more detail in **Chapter 4: Aim, Methodology and Research Questions**, to briefly cover the outputs from each stage.

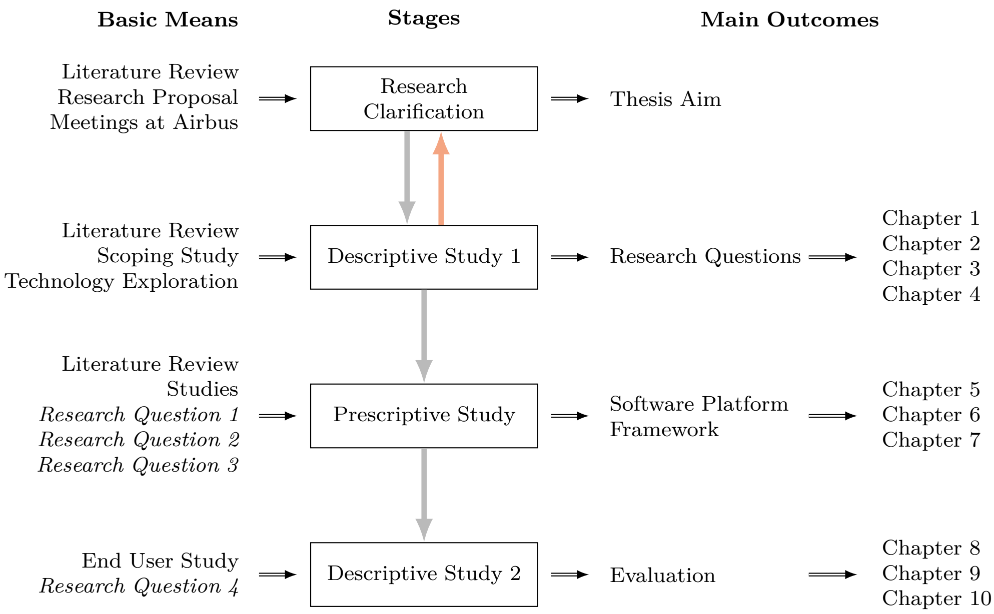


Figure 1.4: The Design Research Methodology Framework as applied in the thesis

The process of achieving the aim and bridging these gaps in knowledge involved a number of steps that are presented throughout this thesis. See Figure 1.5 for the thesis structure. This first step, presented in **Chapter 2: Industrial Context and Scoping Study**, explores and identifies the 'information landscape' at Airbus. The chapter presents an analysis of the search queries that Airbus personnel enter into their Business Search enterprise-wide search engine. A key finding, that Airbus personnel search for product related information, provides the justification for the product centric nature of model-based information navigation.

Chapter 3: Literature Review presents a literature review aimed at better understanding the wider fields surrounding the concept of model-based information navigation. The fields of data, information and knowledge management from both science and in practice at a range of engineering disciplines were explored alongside engineering information management techniques such as product life-cycle/data management systems and building information management systems.

In the methodology chapter (**Chapter 4: Aim, Methodology and Research Questions**) the development of a prototype system, intended for the use in later studies, is presented. The chapter describes a prototype system constructed on Internet technologies including the latest web-based three-dimensional acceleration software libraries WebGL [16] and Three.js [17]. An iterative development cycle and the regular feedback and input from both academics and personnel at Airbus is described.

Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing seeks to determine the appropriate techniques for linking documents to the product structure. Traditional search engines use tried-and-tested techniques for extracting text from documents and producing a search index, essentially, a list of all the terms in every searchable document linked to a list of all the documents containing that term. A key difference between the model-based and traditional text-based approach to search is that one cannot search for 'things' that are not visible in the model-based approach. While at first this may appear to restrict the search process, it also allows for a more intelligent approach to document indexing given that components can be treated like classes in a document classification problem.

Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments explores the processes of actually navigating to the location of information within the three-dimensional environment. When one searches for information using a text-based search system, the results are brought back to the user and presented in a list of results. In a model-based interface, the results are all located in the virtual environment from the offset and the user then must move to the source of information. This is more akin to how one would search for objects in the 'real-world'. To avoid the user 'turning over every stone' to find information, the system must support the navigation process, if the required information is in the centre of a car engine, how can the system aid the user to access that information as quickly as possible? Chapter 6 answers this.

Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation identifies the appropriate techniques for displaying information within the three-dimensional virtual environment. One can think of this as the fruit in a tree: nature evolved fruit such that its colour contrasts with the green leaves of the tree. This results in fruit being relatively easy to identify, then an animal picks it and distributes the seed. In terms of model-based information navigation, the fruits are documents and seeds are the information and knowledge contained within. The system must then provide a means of identifying document such that they can easily be found. Chapter 7 describes the design and testing of a number of visual information objects, markers that are placed in the model-based virtual environment to identify the existence and location of information/documents.

Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design and **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results** combine to answer the fourth and final research question. The Research Questions 1, 2 and 3 focused on the appropriate techniques for developing a model-based approach to information navigation, the fourth research question is aimed at determining the whether a model-based approach to information navigation improves information access and knowledge discovery. Chapter 8 describes the design of an A-B study that compares a model-based and a more traditional text-based approach to search. The two are systems designed such that the user interfaces are the main differences between them, allowing for a more direct comparison of the effect of the model-based interface when compared to a text-based interface. Chapter 9 presents the results from the study.

While chapter 2 contains the main thesis literature review, chapters 5, 6, 7, and 8 and 9 together, each contain a literature review themselves. These four chapters act as stand-alone pieces of work with each presenting a literature review, study, results and contribution. **Chapter 10: Conclusion, Implications for Airbus, and Future Work** then closes off the thesis with a reflection on the thesis and research questions as a whole before considering a range of topics deemed worthy of future work.

1.5 Summary

In their drive to deliver state-of-the-art knowledge management technologies and practices, the Airbus Knowledge Management team set the task of improving engineering

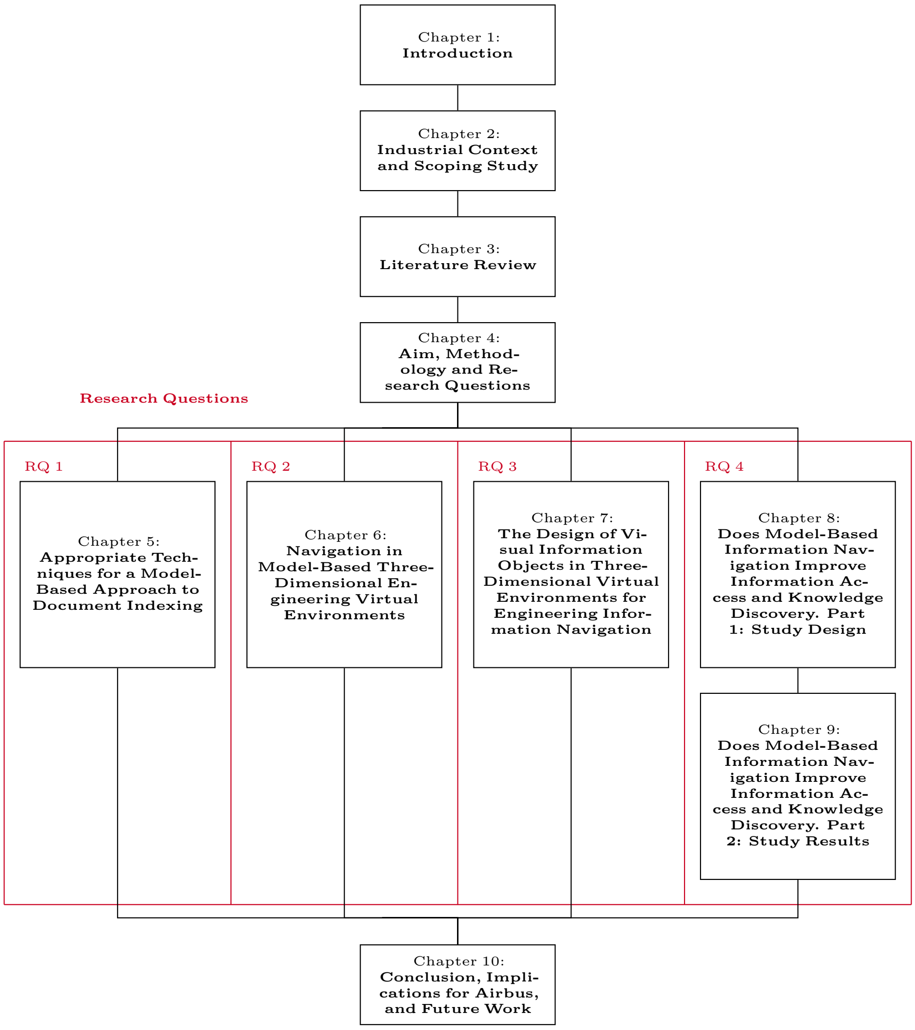


Figure 1.5: The thesis structure

information access and knowledge discovery. A scoping study identified the Airbus 'information landscape' and a literature review showed how context and domain specific implementations can improve the information technology systems provided to engineers. Based on the premise that engineers think visually and functionally, this thesis proposes model-based information navigation as a means of delivering a highly intuitive approach to document search and navigation. This thesis presents a framework for delivering such a solution, covering the appropriate techniques and considerations for building the system, as well as by exploring the affordance of such an approach when compared to a more traditional text-based approach to information navigation.

Chapter 2

Industrial Context and Scoping Study

2.1 Introduction

Airbus is one of the world's leading manufacturers of aircraft. Its heritage began in 1967 when a French and German consortium working with the United Kingdom and the Netherlands began collaborating on the first twin-engine, wide-body passenger jet[1], the Airbus A300. Since then, the organisation has expanded to include a number of European aerospace organisations and through a number of mergers and transfer of assets, Airbus SAS became a company in its own right in the year 2000[2]. With this came the expansion of its product family and the services that it provides. The commercial aeroplane product family includes a range of long and short haul aircraft: the A320, A330, A350 XWB and the A380[3].

Airbus has grown to become a world leader in aircraft design and now competes head-to-head with the likes of Boeing. As such, Airbus is driven to continually push the boundaries of aerospace research and associated disciplines in order to remain competitive. This thesis is a result of this push and is focused on the associated discipline of knowledge management, or how the organisation captures and disseminates its information to best support its engineers.

This chapter outlines the industrial context in which the thesis sits. It begins with

[1] `http://www.aircraft.airbus.com/company/history/the-narrative/early-days-1967-1969/`. Last visited on 06th of November 2017

[2] `http://www.aircraft.airbus.com/company/history/the-narrative/record-breakers-1993-2000/`. Last visited on 06th of November 2017

[3] `http://www.aircraft.airbus.com/aircraftfamilies/`. Last visited on 06th of November 2017

a brief overview of Airbus as a global engineering organisation and is followed by an overview of their products and services that raise challenges in how the organisation operate and remain competitive.

Specifically, this chapter covers:

- The current knowledge management strategies developed by Airbus to support the organisation
- A description of distributed nature of Airbus engineers
- A description of the nature of Airbus engineering artefacts
- A study into the search practices at Airbus, i.e. how Airbus personnel search for information and types of information that they are searching for
- The selection of a number of Airbus functions with which to align the thesis such that they can act as the basis of use-cases

The purpose of this scoping study is to determine and identify:

- The relationship between Airbus personnel and their data
- The domain specific features of Airbus, their products and their personnel such that the features can be leveraged to improve information access and knowledge discovery
- A starting point for the literature review covered in **Chapter 3: Literature Review**
- Airbus function(s) with which to align this thesis

2.2 Airbus

This section gives a brief overview of the Airbus Group to provide an underpinning to the industrial context of this thesis. It will cover Airbus as a global organisation before discussing the challenges associated with designing, manufacturing, selling and servicing aircraft. Finally, the Airbus Knowledge Management team is briefly introduced before a more complete description in **Chapter 3: Literature Review**.

Figure 2.1: Airbus Locations across the world[10].

2.2.1 A Global Engineering Organisation

Airbus is a world leading designer and manufacturer of aerospace products and services with its head quarters based in Toulouse, France[4]. The organisation is split into three sectors: commercial aircraft, helicopters, and defence and space. The support for this thesis is from the commercial aircraft sector and specifically the Knowledge Management team. The customer aircraft sector being the design, manufacture, sales and service of aircraft that are typically brought by airlines from across the world.

Airbus is a global organisation and operates with a staff of over 130,000[5] in sites across North America, Africa and the Middle East, Asia and its home region of Europe[6]. Design Offices are situated in Europe, North America, India and China[7]. Production sites are situated across Europe [8], China and the United States of America[9]. Spare part centres operate out of Hamburg, Frankfurt, Washington DC, Dubai, Beijing and Singapore. Training centres operate in Toulouse, Miami, Hamburg and Beijing and as part of In-Service support they operate over 150 field service offices globally. Figure 2.1 shows the global extent of the Airbus organisation. This truly global structure results in staff being distributed across multiple locations and organisational functions spread across multiple sites and countries.

With any large organisation, communication is a part of the day to day operation.

[4] `http://www.airbus.com/company.html`. Last visited on the 6th of November 2017

[5] `http://company.airbus.com/company/about-airbus.html`. Last visited 07th of November 2017

[6] `http://company.airbus.com/company/worldwide-presence.html`. Last visited on 06th of November 2017.

[7] `http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/design-offices-and-engineering-centres/`. Last visited 07th of November 2017.

[8] `http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/production/`. Last visited 07th of November 2017.

[9] `http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/transport-of-major-aircraft-sections/`. Last visited 07th of November 2017.

[10] `http://www.aircraft.airbus.com/company/`. Last visited 07th of November 2017.

Whether through and email, a video conference, a formal report or sharing specific document types such as CAD drawings or source code, organisations will struggle to operate without the flow on information between staff. The global nature of Airbus compounds this further, with the additional challenges of staff working in different time zones, within different legal and regulatory systems and with staff speaking a range of different first languages. This is part of the reason why knowledge management activities and systems are important as they are intended to ease the flow of information.

2.2.2 Products and Services

The commercial arm at Airbus is divided into both aircraft and services [11]. Commercial aircraft include the single-aisle A320, the wide-body A330, the state of the art A350 XWB and the world's largest passenger aircraft, the double-decked, wide-body A380. In terms of services, Airbus provide 24/7 worldwide maintenance and engineering, upgrades to existing stock ranging from appearance to flight performance, flight operations including air traffic management and crew planning and a global network of training facilities. The nature of these products, regulation and the manner in which airlines and airports operate result in a set of unique challenges that are unique to both Airbus and organisations offering similar products and services.

Challenges regarding Airbus products derive from the fact that aircraft are long-life, high-value, low-volume and safety critical. Long-life artefacts are in-service for many decades and as such information relating to their design, modification, repair and upgrade must be captured and accessible for the duration of potentially half a century or more. High-value means that every effort must be made to deliver artefacts that are right first time as the costs of errors or low quality are substantial and potentially seriously damaging to the manufacturer. Low-volume artefacts mean that there are relatively few products in service when compare to other engineering products such as cars or mobile phones. Safety-critical products simply means that malfunctions could result in loss of life and a range of legislation exists across the world that aim to maximise aircraft safety and minimise the risks to people. To demonstrate all this, during testing the Airbus A380 is put through the equivalent of two and a half times the number of flights expected in 25 years of operations[12], the 2017 average list price for the A380 is 436.9 Million USD[13], as of October 2017, the Airbus Group sold 317 A380 since

[11] `http://www.airbus.com/aircraft.html`. Last visited 10th of November 2017.

[12] `http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/test-programme-and-certification/`. Last visited 10th of November 2017.

[13] `http://www.airbus.com/content/dam/corporate-topics/publications/backgrounders/`

its launch[14] and the process of certification by the European Aviation Safety Agency (EASA) and the United States of America's Federal Aviation Administration (FAA) took over five years and encompassed 2600 flight hours with five test aircraft[15]. Aircraft then must be manufactured and maintained to a high standard but the low-volume of in service aircraft results in there being few chances to gather and learn from in-service data that could be used to improve Airbus products through better design, performance and preventative maintenance.

The service that Airbus offer that is the main focus of this thesis is In-Service Maintenance and Engineering, the activities of which are explained in more detail in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**. Airlines buy aircraft and make money by selling flights on those aircraft, as such, if an aircraft is unable to fly then it is simply unable to make money. There are also additional financial costs incurred if an aircraft is left stranded at an airport or misses its designated landing/take off time slots[16][17]. It is then paramount that aircraft are operational for the maximum amount of time and organisations that provide in-service maintenance support must achieve this if it is to make a success of selling such a service.

The nature of these product and services means that maximising the dissemination, use and reuse of information is desirable. An example of this is being able to capture and learn from every maintenance repair case and disseminate that information across fleets. This could lead to preventative maintenance scheduled as part of routine maintenance and reduce the cases of aircraft being stranded at airports. As a result, Airbus has dedicated team who look specifically and the capture and dissemination of information across the organisation, the Airbus Knowledge Management team.

2.2.3 Knowledge Management

The Airbus Knowledge Management team are a group of individuals situated at sites across Europe tasked with the capture and dissemination of Airbus knowledge and in-

`Backgrounder-Airbus-Commercial-Aircraf-price-list-EN.pdf`. Last visited 10th of November 2017.

[14]`http://www.aircraft.airbus.com/aircraftfamilies/passengeraircraft/a380family/`. Last visited 07th of November 2017.

[15]`http://www.aircraft.airbus.com/market/how-is-an-aircraft-built/test-programme-and-certification/`. Last visited 10th of November 2017.

[16]`https://www.usatoday.com/story/travel/flights/todayinthesky/2016/12/15/dot-tarmac-delay-fine-american-airlines/95464960/`. Last visited 10th of November 2017.

[17]`https://www.usatoday.com/story/travel/flights/todayinthesky/2017/09/18/frontier-airlines-fined-1-5-million-long-tarmac-delays-denver/676175001/`. Last visited 10th of November 2017.

formation. This thesis is sponsored by the Airbus Knowledge Management team and is part of their wider activities of improving the capture and dissemination of information across the organisation. Their knowledge management processes are based on the premise that an organisation's most valuable resource is the knowledge of its people. As such the team engages and encourages people to create, share and use knowledge for their own benefit and for the benefit of the organisation.

Over the last two decades Airbus has successfully implemented a number knowledge management solutions which have been in use for the design and development of the A380 and A350 programmes. These activities are centred around their Knowledge Management Wheel, see Figure 2.2. The wheel is divided into eight segments representing the knowledge management portfolio. These cover Business Search, Lessons Learned Reports, Innovation Management, Connect@Airbus (a social media application that doubles as a Yellow Pages of business experts), Professional Networks, Expertise Transfer (to minimise the knowledge loss when staff change job), knowledge management Diagnosis (auditing and delivering road maps for knowledge management practices) and Knowledge Capture and Publishing (generating 'knowledge books' to pass the correct knowledge to complete tasks). These are, in essence, a set of procedures and software systems designed to solve a particular information need (Yellow Pages of business experts for example) or to combat a source of information loss (someone leaving the company or changing jobs). The knowledge management team are engaged in a drive to innovate and improve the services that they offer and this thesis is an example of this innovation. A more complete picture of the Airbus Knowledge Management Wheel is given **Chapter 3: Literature Review**, alongside knowledge management activities from other engineering sectors.

The Airbus Knowledge Management team strive to deliver world-leading, state-of-the-art knowledge management technologies and strategies and as such fund and participate in a range of research projects. The work presented in this thesis being one of a number of examples [18][19][20][21][22].

2.3 Remit

Airbus is a global engineering company that faces a number of challenges around the nature of its products, services and organisational structure. Airbus products are long-life, high-value, low-volume and safety critical and Airbus must achieve safe products that are worthy of a high value if it is to compete in a global market. The relatively

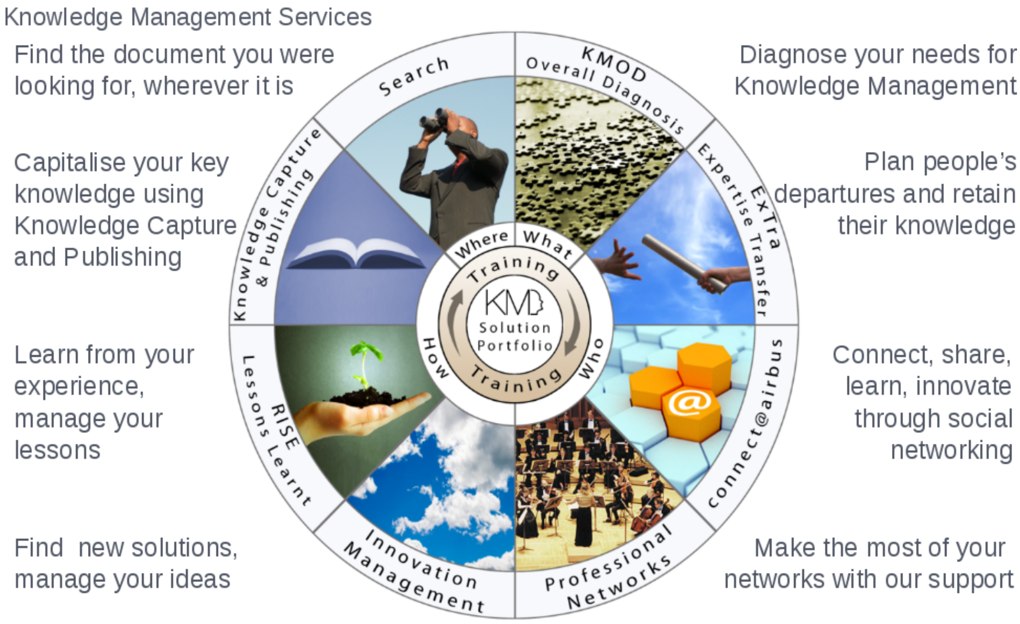


Figure 2.2: The Airbus Knowledge Management Wheel

low volume of aircraft in service means there are few options to learn from real world performance and this is a problem identified by the organisation. In addition, one of the services offered by Airbus, in-service maintenance and engineering, can result in high costs to the customer airlines if Airbus provides poor service and aircraft are left grounded. One method in which Airbus aim to mitigate challenges such as these is to maximise the organisations capture and dissemination of information and knowledge and it is this desire that lead to the proposal behind the work presented in this thesis.

The remit of the research presented in this thesis was written by the members of the Airbus Knowledge Management team and researchers at the University of Bristol. This proposal called for innovation in the field of knowledge management, an audit the Airbus information landscape, the identification of and alignment with Airbus function and the development of innovative solutions based on the fields machine learning and information/knowledge visualisation. The goal of this audit then was to narrow down this broad research space while maintaining a connection to Airbus. The information audit took the form of a scoping study which is detailed in the remainder of this chapter.

2.4 Scoping Study

Airbus fund many research projects that do not necessarily have an impact on day-to-day operations that the organisation envisages during the early phases of the project. From the offset it was felt that this project and the stakeholders involved provided a better than usual opportunity to enable and support change within specific functions within the organisation and Airbus were keen to capitalise on this. However, Airbus were also acutely aware of the inherent difficulties in gaining access to functions, personnel, and data and as such, an 'information audit' objective was incorporated into the project plan.

The purpose of the audit was to identify applicable and promising data sets and functions that both align with the research area while being able to support research activities. Expanding on this remit, the audit was also used to elicit and characterise the relationship between Airbus personnel and Airbus data, specifically, how personnel search for and access information. Information that was deemed useful in shaping and underpinning the work presented in this thesis: a better understanding how Airbus currently uses information leading to research output with better alignment. Correspondingly, this chapter explains the audit and the findings which ultimately led to the research questions presented in **Chapter 4: Aim, Methodology and Research Questions** and the use-cases outlined in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**.

2.4.1 Methodology

Over the first 18 months of the project at Airbus sites in Bremen, Toulouse, Madrid and Filton, the 'information audit' involved a combination of meetings and exploring data sets from a range of Airbus functions. These functions included Knowledge Management's Business Search and Lessons Learned, the Skills and Competency team, and the Wing In-Service teams. The personnel involved in these meeting covered a range of end users, including management, engineers and support.

The purpose of these meetings was to ascertain:

- The types of information being used
- The means by which information was being access
- How the teams/data fitted into the wider knowledge management strategy
- Whether there was scope for research
- Whether the function could support research activities over the length of the project

Ultimately, the alignment required a function with the data sets, use-cases, and resource to support the project over an appropriate time line. During the audit of these functions, the details relevant to the scope of this thesis were also captured - effectively any details relating to the access and use of knowledge and information.

2.4.2 Results

At the time of the study (2016-2017) the Skills and Competency and the Lessons Learned functions were the subjects of internal change programmes that gave rise to a level of uncertainty in both the structure of their data, and they ability to support research activities going forward. While they both embodied technical information relevant to the scope of this thesis they were deemed unsuitable for use. Further details of these two functions are not relevant to the wider thesis and as such are not included. The Business Search and Wing In-Service systems were however selected for further detailed investigation.

Information Landscape

While the Business Search function was engaged and willing to support research, the organisation's dependency on search and the internal infrastructure resulted in a high probability of not obtaining permission to both experiment with the system and implement changes. This function was however the most useful in understanding the information needs at Airbus as the service offers high level search capability of a range of systems across the organisation.

Airbus holds vast volumes of data across multiple systems tailored to specific departments and functions. Table 2.1 is an example of this, and shows a snapshot of a number of the current search systems used within the organisation alongside the number of data sources indexed, the number of documents contained within it, and the size of each index.

System	Number of Data Sources	Number of Documents Indexed (Millions)	Size of Index (GB)
Business Search	16	5	107
One Search for All	19	1.3	67
AirbusWorld	36	37	634
AirbusSupply	5	0.5	41
Enterprise Search for Applications	4	2.1	89
SFS Search	1	116	7000

Table 2.1: A snapshot of a number of the search systems used within the Airbus Group

In addition to these systems and the indexed data sources, Airbus also utilise a 5 TB SharePoint farm, a 200 TB SAP environment with over 250 applications, PLM systems with approximately 600 applications, ICT support for over 400 applications and their archiving system contains around 140 million documents. The information landscape at Airbus is consequently one of large quantities of data spread over distributed technologies with tailored search engines constructed around specific business requirements.

Within these data sources, data types range from highly structured databases through to semi-structured templates and forms, to unstructured free-text documents. Table 2.2 shows examples of the types of data utilised at Airbus. These data types include unstructured, semi-structured and mixed media documents stored within a range of systems.

Name	Description
Skills and Competency Catalogue	A database of short text descriptions
RISE (Airbus Lessons Learned)	Semi-structured forms stored in a SAP environment
Airbus In-Service maintenance reports	Mixed-media PDF documents containing text (both handwritten and typed), technical drawings and photographs

Table 2.2: Examples of the range of data types used at Airbus

The Business Search team also provided a list of the top 500 search queries collected from 1 January 2014 through to 30 June 2014. This data covered nearly 1.1 million searches with approximately a third of those being unique and executed by more than 68,000 unique users. This contribution to the thesis was presented in the paper "*Improving Enterprise Wide Search in Large Engineering Multinationals: A Linguistic Comparison of the Structures of Internet-Search and Enterprise-Search Queries*"[23]. This paper firstly compared the Airbus top 500 queries with those gathered by the internet keyword research organisation WordTracker.com. WordTracker provided the top 500 queries for January 2015 which allowed for the direct comparison between how users search the internet and how Airbus users search within the Business Search system. Tables 2.3 and 2.4 show the top 10 queries from the two top 500 lists.

The comparison comprised of a Part-Of-Speech (POS) analysis of the two lists and examining the word class (noun, verb, etc.) frequencies. The paper concludes with the findings that Business Search users within Airbus are more likely to phrase their queries using nouns, with 97% of search queries containing nouns compared to 89% for Internet queries. Airbus Business Search users also use far less variety in the formulation of queries, with queries falling into 41 word classes with just four of those required to cover 80% of queries, compared to 94 for Internet and 51 to cover 80% of queries. The

Internet Search Queries	
Query	Frequency
youtube	9924821
movies	8721604
facebook	8085544
google	6968440
entertainment	6067158
search	5186360
craigslist	4888389
kinox	4828994
hood stars clothing	3735957
download	3006655

Table 2.3: The top 10 Internet search queries

Business Search Queries	
Query	Frequency
docmaster	8736
icc	7186
lexinet	7022
webex	7012
pwinit	6591
uvisit	3982
airnav	3310
eds	2967
zamiz	2766
edms	2692

Table 2.4: The top 10 Business Search queries

Class	Frequency
Application	172
Document	108
Activity/Process	99
Organisation	81
Product	80
Project	25
Role	23
Devices	19
Discipline	17
Gate	2
Member	2
Unknown	80

Table 2.5: The classified top Business Search queries

key finding was that Airbus Business Search users are more likely to search for specific ‘things’.

The paper then explored this further by asking Airbus to group the top 500 terms into a range of classes relating to the context of the search (i.e. the search was about a product/person/etc.), The classes were influenced by published work [24] performed at EADS (European Aeronautic Defence and Space) formally the parent company of Airbus (has now been re-branded as the Airbus Group). Table 2.5 shows the results of this analysis. Of the total number of Business Search queries and other than the 80 classed as Unknown, the highest top 5 classes were *applications*, *documents*, *activities/processes*, *organisation* and *product*. These classes cover 78% of business search queries.

Combined, all these findings show that Airbus users are more likely to search for real-world ‘things’ and that those ‘things’ are easily classified into the day-to-day business activities. This forms the foundation of this thesis in that:

1. Traditional search systems are largely based on Internet search technology but

there is a clear differentiation between Internet and Airbus search queries.

2. The structured nature of the classes in Table 2.5 can provide context and structure to search systems that is not known to be used within Internet search engines.

An argument could be made to design and test a system that exploits the structure from the full list of classes, however the decision was made to focus on a single class and to study that in depth rather than produce a high-level multi-faceted approach. Given the state-of-the-art of engineering data management, PLM and PDM systems aim to place the product at the heart of the data management strategy. In accordance with this, the *product* class is considered in this thesis.

Function Alignment

As a large multinational organisation, Airbus divides its organisational structure into hierarchical functions tasked with specific activities that support different aspects of the aircraft and phases of its life cycle e.g. wing development, flight physics and knowledge management. One branch of the organisational structure is the In-Service Repair and Maintenance function located at sites across the world, including at the Filton base in Bristol, UK. The In-Service function comprises a number of contact centres and specialist teams of engineers that work in response to airline maintenance teams that are in need of repair advice and support.

As with any engineering artefact, aircraft require continuing maintenance and repairs to remain operational. Aircraft face the additional challenge of strict regulations that govern airworthiness [25]. Any maintenance and repairs performed on aircraft must meet the highest of standards for that aircraft to be legally allowed to fly and carry passengers. In an attempt to aid airline maintenance teams, Airbus provides the Service Repair Manual (SRM), a detailed set of instructions, guidelines and tolerances on many types of damage and repairs. Repairs performed in accordance to the SRM do not require any additional checks for the airline to meet its safety commitments.

Over the life-cycle of an aircraft it is inevitable that damage will occur that is not covered by the SRM or that falls outside the tolerances outlined in the SRM. In these cases, airlines contact an In-Service Repair and Maintenance team responsible for the part of the aircraft in question and this team of engineers evaluate damage and respond with instructions for an appropriate repair that is approved against the airworthiness standards. Designing these repairs can range from a straightforward replacement of a part to needing to perform full stress or aerodynamic analysis on a proposed fix.

Airlines are imposed with large fines when aircraft fail to adhere to flight schedules, therefore any aircraft that are not able to fly are costing rather than generating it. Every fault that prevents an aircraft from operating can be expensive and as a result airlines demand a quick response when they report a fault. In response, Airbus endeavours to meet these expectations and the re-use of past repair cases is one activity that helps achieve this. If a similar incident has occurred in the past, the approved repair can be re-used or adapted. This removes the need to perform the analysis required to meet the airworthiness standards and so reduces the response time.

The high-cost, low-volume and long-life nature of Airbus aircraft results in a situation where there are relatively few opportunities to learn and apply knowledge from previous incidents. The A320, for example, has been in service since 1984[18] and the new A380 has sold just 317 since its launch in 2007[19]. There is then a need to maximise the reuse of knowledge learned during every repair. In response to this, the In-Service team responsible for the A320 wings maintains a database of past cases and the first step in every new reported incident is to search that database for a past case whose repair can be applied to the new case. The current system is called Daedalus and is an MS Excel spreadsheet with an ontological search implemented [8].

After meeting the Wing In-Service team at Filton, examining their data sets and operational procedures, this function was chosen as the most appropriate to align with. Effectively, the Wing In-Service team was chosen as the basis for real-world use-cases which frame the two of the research question studies. The details of these use-cases are presented in **Chapter 4: Aim, Methodology and Research Questions**. The primary reasons for aligning with the Wing In-Service team were:

- The repair requests are tied directly to the product structure and as such aligns with the product structure query class identified in Table 2.5, Subsection 2.4.2 of this chapter
- The team generate a range of document types covering a range of different topics, giving a rich range of data types
- Document searches are performed as part of every repair request and so the team are familiar with search and depend on search

[18] `http://www.aircraft.airbus.com/company/history/the-narrative/fly-by-wire-1980-1987/`. Last visited 21st of November 2017.

[19] `http://www.aircraft.airbus.com/aircraftfamilies/passengeraircraft/a380family/`. Last visited 21st of November 2017.

- The team works with a corpus of engineering-related documents that relate to the aircraft systems, sub-systems and parts
- The 2013 repair reports for the single-aisle A320 wings (240 in total) were provided for analysis
- The time critical nature of the repair request makes time a critical measure for evaluation meaning any improvement to the process will have a significant benefit
- The team have supported research activities in the past and so are familiar with the process and level of input entailed

The information needs of the In-Service function are discussed in detail in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**. A second audit was performed on the reports used within the function and from this a set of tasks were generated that were ultimately used to generate a set of use-cases and the testing strategy presented in **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**.

2.5 Discussion and Conclusion

This chapter outlines Airbus's position as a global engineering organisation operating in a competitive market place. In order to remain competitive Airbus both invests in knowledge management and in research activities such as this thesis, with the goal of maximising the retention and re-use of information across the organisation.

Keen to maximise the output from the work presented here, the initial research proposal called for a scoping study to better understand the information landscape at Airbus and to align with Airbus functions that would support research activities. Through understanding the information landscape, the Business Search system was highlighted as a cornerstone to knowledge management activities and deemed appropriate as an area that would benefit from improvement.

This chapter the set out to determine and identify:

- The relationship between Airbus personnel and their data
- The domain-specific features of Airbus, their products and their personnel such that the features can be leveraged to improve information access and knowledge discovery

- A starting point for the literature review covered in **Chapter 3: Literature Review**
- An Airbus function with which to align this thesis

The outputs from this chapter are an understanding that:

- The Airbus 'information landscape' is one of multiple dispersed systems tailored to specific departments and functions
- Airbus personnel search queries can be classified into a number context classes, the product being the one of interest to this thesis
- Airbus products face a number of challenges when compared to other engineering products given aircraft are long-life, high-value, low-volume, and safety critical
- The re-use of product related information is important but difficult and this in emphasised in the Wing In-Service team, the function with which this thesis as been aligned
- The fields of engineering knowledge management and engineering information management are the primary focus of the literature review

The final item in this list requires explanation as it is not covered in the main body of the study. Given the Knowledge Management team set the remit for this thesis, it is understandable that the wider knowledge management field be explored. The topics discussed in this chapter and discovered as part of the scoping study also relate to the wider field of engineering information management. Systems such as product lifecycle/data management and building information management become relevant as they are tailor-made engineering solutions to managing engineering information. These are now discussed in the next chapter, **Chapter 3: Literature Review**.

Chapter 3

Literature Review

3.1 Introduction

It is widely accepted that an organisation's greatest asset is the data, information and knowledge captured within an organisation's databases, documents and people [4]. As such, the efficient dissemination of knowledge across product life cycles is something highly desired at the Airbus Group and the Knowledge Management team has grown to fund, implement, and apply the latest research and approaches to best achieve the efficient capture, storage and dissemination of knowledge. The very nature of Airbus' products (long-life, high-value, low-volume and safety-critical) means there are relatively few opportunities to learn, improve and evolve the product during its life-cycle. With this in mind, it is vitally important that the maximum value is realised from every piece of information and every aspect of the organisation's knowledge in order for Airbus to maintain its market position.

The journey to efficient knowledge management begins with the understanding of what are data, information, and knowledge. Once this is understood, the strategies to capture, store and disseminate them can be explored. From this point, engineering information management has evolved to generate domain-specific approaches that tailor information management solutions to the product and product life-cycle. These solutions range from data management strategies on file servers to incorporating information within the CAD models themselves. This chapter follows this journey, exploring the foundations and state-of-the-art science in each field.

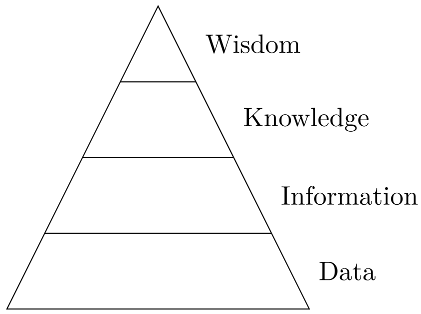


Figure 3.1: The DIKW hierarchy

3.1.1 Data, Information, and Knowledge

The relationship between Data, Information, Knowledge and Wisdom (DIKW) is a widely debated topic and this thesis is not concerned with adding to the debate. As such and for the purpose of this thesis the traditional DIKW hierarchy is used [26]. This framework describes information in terms of a collection of data, knowledge as a collection of information and wisdom as a collection of knowledge. See Figure 3.1.

It is, however, appropriate to define data, information, and knowledge for use within the context of this thesis and remainder of this subsection does this from engineering domain. Rowely [5], presents a literature review of DIKW hierarchy bringing together definitions from a range of information systems and knowledge management textbooks. To ensure alignment with literature, Rowely's collated definitions are used as a basis for those presented.

Data is raw, unstructured and unprocessed with no context (and so no meaning) and describes object properties, e.g. digits, letters, words/keywords, etc. In terms of engineering data, these could be the raw numerical production scores alongside dates and times and stored within a database. Other examples include the keywords and queries formed by search engine users. These are letters/numbers and words that are given full context once included in the documents returned by the search engine.

Information is data that is formatted for the purpose of conveying purpose. Graphs (line, bar, pie, etc.), reports, CAD drawings, etc. Displaying production scores against time in a histogram conveys the measure of performance in a meaningful and useful way. Other examples include assembly instructions, where the physical structure of an artefacts components are visually displayed along with the details on how they fit together.

Knowledge is learned information and/or data such that it is applicable. So if the assembly instructions contain information, the act of reading, understanding and applying that information is knowledge. Knowledge is also split into tacit and explicit.

- Tacit knowledge is the personal knowledge of an individual and includes insights and intuitions however it is difficult to capture and share Speaking a foreign language in one such example
- Explicit knowledge is documented tacit knowledge that has been codified and documented, textbooks and thesauri are examples of these

Duffy states the difference between information and knowledge as information being cheap and easy to obtain while knowledge is information enriched through people's interpretation and analysis. One question then is how valuable is one piece of information or knowledge over another. Zhao et al. aim to answer this question [27] through a technique that categorises engineering information through seven characteristics: accessibility, usability, currency, context, accuracy, availability, and relevance. Zhao et al. do not formally define these attributes within their publications, however, see Table 3.1 for the list of attributes and the description for each of these used within this thesis. These descriptions are based on the Oxford English Dictionary[1] definitions and deemed appropriate to the field of information and information access.

These characteristics are utilised within a probabilistic Bayesian network that predicts the value of information. So the journey from data to wisdom can be seen as data being captured and converted into valuable information, which is then interpreted to from knowledge.

It could be said that the more knowledgeable and wise a workforce is, the better they perform and hence the better the organisation performs. Capturing data, information and knowledge and converting it into knowledge that is made available to staff is then beneficial and this is where knowledge management comes into play. This chapter now moves on to discuss knowledge management from both a literature and 'real-world' perspective.

3.2 Knowledge Management in Literature

The knowledge management ethos is based on the idea that one of an organisation's greatest assets is the knowledge contained within it and it is good business practice to not only

[1] `https://en.oxforddictionaries.com.Visited2018-12-17.`

Character	Description
Accessibility	OED definition and example: The quality of being able to be reached or entered. 'We focus on issues that address the needs of bus users, such as bus stop accessibility and the comfort and safety of buses.' In terms of this thesis, this definition translate into the users' ability to access a piece of information. E.g. '*I know the information is in the system but how do I find it?*'
Usability	OED definition and example: The degree to which something is able or fit to be used. 'As for usability, it is extremely easy to use, with a lot of energy put into sensible search functions.' While it is a characteristic of information itself (*how easy is it to use this piece of information?*), in terms of this thesis, usability refers to the ease of use of the system to support the access of information.
Currency	OED definition and example: The time during which something is in use or operation. 'no claim had been made during the currency of the policy.' In terms information, this can relate to the relevance of information to a particular time. In terms of information access and this thesis, currency is the time taken to access a piece of information.
Context	OED definition and example: The circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood. 'the proposals need to be considered in the context of new European directives.' In terms of this thesis, context is the relationship between information and the information need e.g. an engineer searching for the term 'cat' is more likely to be searching for a catalytic converter rather than the animal.
Accuracy	OED definition and example: The quality or state of being correct or precise. 'In these cases, the likely outcomes can be predicted with some accuracy.' In terms of this thesis, accuracy is how well a piece of information meets an information need. E.g. '*I'm looking for the recent profit/loss reports but all I'm getting is payroll, this system isn't very accurate.*'.
Availability	OED definition and example: The quality of being able to be used or obtained. 'Remove bird feeders to help lower the availability of food for voles.' Within the context of this thesis, availability is whether a piece of information is available for access or not. E.g. '*I know the information exists but the system isn't able to access it*'.
Relevance	OED definition and example: The quality or state of being closely connected or appropriate. 'Having a better understanding of the make-up of my students allows me to establish the relevance of the course material to their specific interests.' In terms of this thesis, relevance is a measure of appropriateness of a piece of information to an information need, e.g. '*I searched for profit/loss reports but all I'm getting is payroll reports, payroll reports aren't relevant*'.

Table 3.1: Seven characteristics of information and the corresponding descriptions used in this thesis.

safeguard knowledge, but to maximise its re-use throughout the organisation [4][28]. As a result, over the last few decades many organisations have adopted a range of practices aimed at the capture, storage and dissemination of knowledge [29][30][31][32][33]. Airbus in no different [34][35]. **Chapter 2: Industrial Context and Scoping Study** introduced the Airbus 'Knowledge Wheel' and the knowledge management activities with which Airbus is engaged. This is explored further later in this chapter.

The early foundations of knowledge management are somewhat vague, however a review performed by Alavi and Leidner [30] states that the work is built upon the work of Penrose in 1950[36]. In their review written in 2001, Alavi and Leidner give a complete picture of the factors surrounding knowledge management that are still relevant today. Their paper begins with an overview of data, information and knowledge, similarly to Section 3.1.1 of this chapter. This is followed by a discussion of knowledge management in organisations, including some reasons behind the practice e.g. the loss of expertise through staff leaving and the challenges of finding and using an organisation's existing knowledge. Alavi and Leidner then introduce a framework for the analysis of information systems to support knowledge management.

The Alavi and Leidner framework discusses knowledge management in terms of the creation, storage/retrieval, transfer, application of knowledge. This thesis however follows other work in discussing the practice in terms of the capture, storage, and dissemination [4][37]. The primary difference in the framework used in this thesis assume knowledge already exist and as such is not concerned with its creation. Airbus as an organisation creates and uses knowledge and has vast repositories of historical knowledge (see **Chapter 2: Industrial Context and Scoping Study**). It is then the improved capture, storage and dissemination which is of concern to this thesis.

3.2.1 Capture

Knowledge capture is the collection of the organisation's knowledge. It exists in two forms: the first is where, through a range of activities, knowledge is actively sort and documented. The second is through the management of knowledge documented through the day-to-day work practices. Examples of each of these can be seen in the Airbus Knowledge Wheel 3.2. The Knowledge Capture and Publishing and Expertise Transfer are examples of this, where knowledge experts are brought into a department and facilitate the transfer of knowledge from staff to either 'Knowledge Books' or to new recruits in the case where a member of the team is leaving. Search is an example of where existing knowledge is captured made available.

Many of these activities involve extracting elements of documents and structuring them such that they can be stored and made available for dissemination. Examples from the active approach to knowledge management include the creation of Wikis [38] (online repositories of knowledge) or Yellow Pages [33] (contact directories). Structuring existing document collections is a larger problem given these forms of documents (reports, spreadsheets, etc.) are often unstructured free text [39]. Examples here require pre-processing to generate that structure, often involving information retrieval and linguistics to exact the useful and meaningful knowledge [40]. The development of a context-based search platform for example [41]. The process of creating search indexes is discussed further in **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**.

Within the engineering field, there are a number of knowledge management research areas that receive focus more than others. Design rationale capture, design process modelling, design support systems, and structured document retrieval to name a few.

Design Rationale are the decisions made during the design process. The justification for the artefact form, material, etc. To look at an example, if during the service phase of the product's life-cycle a particular component develops a defect and needs a re-design, the process is far more complex if the rationale behind the original design is not understood. As such, its capture for re-use should lead to cost and efficiency savings. DRed (Design Rationale editor) [42] is a tool specifically designed for the capture of design rationale. Its purpose is to allow engineers to diagnose problems from their initial understanding through to designing the solution. Information is captured throughout and as such is available for dissemination.

Design Process Modelling is the process of effectively recreating the stages in design via the analysis of the documents created throughout design activities. Lan et al. [43] have shown how the machine learning techniques of deep neural networks and process mining, meta-data extraction, and classification can re-create design stages and provide design patterns that can support future design activities.

Design Support Systems is a term used here to cover a range of research that aims to support engineers during the design process by providing them with access to information and knowledge. Howard et al. discussed information as an input to the creative process [44], and identified two distinct information sources: internal (memory) and external (software systems). The paper proposes techniques to improve these two sources as cognitive tools to aid lateral and diverse scanning of memories and new tools for the capture, storage and search of relevant information.

One such example of this shows the use of manufacturing data in the design process [45]. In this work, data mining techniques (decision tree induction and artificial neural networks) were employed in an attempt to re-create and capture the knowledge typically extracted from domain experts. The approach was able to predict machine performance based on early system characteristics and as such was envisioned as having potential in supporting engineers during the design process.

Structured Document Retrieval is the process of extracting useful pieces of information from documents themselves [46]. This area of research aims to provide the engineer with specific 'snippets' of information rather than entire documents and ultimately increase the speed by which information is obtained, such that the engineer need not scan an entire document for the one pertinent piece of information. In a review of structured document retrieval, Liu et al. [46] reported that the approach can make improvements to document classification, indexing, presentation, and querying and ranking.

In a similar vein, Liu et al. [47] also discuss the retrieval of 'document fragments'. The article presents a framework that integrates document structures, make-up technology, automated fragment technology, faceted classification, and document navigation management. The result being a retrieval system that allows users to retrieve information in varying degrees of fragment granularity as opposed to entire documents.

Within the context of this thesis, knowledge capture is deemed out of scope given the remit of the thesis was focused on the knowledge dissemination rather than capture. These topics are covered here in a bid to provide the wider context to the knowledge management problem and better understand the efforts of the Airbus Knowledge Management team.

3.2.2 Storage

The storage of knowledge is largely (but not exclusively) reliant on some form of knowledge management system - information technology designed for the task of storage of digital forms of captured knowledge [48]. The Airbus 'Knowledge Books' are an example of a non-digital media. The digital media however take the forms of word processed documents, spreadsheets, CAD files, source code, etc. all stored in systems such as database entries, stored search engine indexes, intranet based content management systems, building information, product life-cycle and product data management systems, and traditional file storage servers and achieves [49][33][50]. Product life-cycle and data management systems are discussed in greater detail in Section 3.4 of this chapter.

The digital knowledge storage categories [49] relevant to this thesis fall into three categories, systems that allow the user to navigate a document collection, systems that allow you to search the collection and systems that do both. Shared directory on file servers rely on users to manage folder structures such that documents are stored logically and as such can be navigated logically. Intranet based content management systems are an example of mixed approaches to storage. SharePoint[2] being the Microsoft offering. These are a mix of file storage and Wikis where functions are given a 'web page' on an Intranet that contains a mixture of information/knowledge and links to related documents. Search engines then allow users to search for documents using keywords. Document collections are indexed (**Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**) and the index stored as a list of term and corresponding document locations. During a search it is the index that is searched and a list of matching document locations returned [51]. Search is a part of modern life although it does lack the context that a directory or content management system provides.

Some of the most seminal works in the field of knowledge management are those of Duffy from 2000 to 2001. In these articles, [52][53][54][55], Duffy discusses the fundamentals of knowledge management as well as the tools and technologies required to deliver it. From user interfaces being intuitive, responsive, and valuable, to mapping knowledge, data warehousing for structured data, and content management software for unstructured data. The articles cover everything an "information professional should know". While these works are nearing 20 years old and the world has seen technology moving forwards in leaps and bounds, it is interesting that the fundamentals covered in these works are still highly relevant and central to the knowledge management systems required to store and disseminate knowledge.

There are, however, problems with the implementation of these systems. Hicks et al. [2] examined information management across a number of engineering SMEs and discovered a list of 18 issues and from those nine core issues (see Table 3.2) in implementing information management. These nine issues show that while in theory the world of storage in knowledge management is relatively straight forward, the practical 'real-world' engineering organisation is a myriad of legacy, distributed, mixed-media, vendor specific, and 'cobbled together' systems that often rely on people to bridge the gap between functionality.

This is not too dissimilar to the Information Landscape at Airbus presented in **Chapter 2: Industrial Context and Scoping Study**. Airbus contains numerous systems, provided by a range of vendors, and distributed across a number of locations. This thesis is however confined by the original remit of information access and knowledge discovery and the storage of knowledge is once again deemed out of scope and only included here to complete the picture of knowledge management considerations.

[2] `https://products.office.com/en-us/sharepoint/collaboration`

Issue
Automatic information exchange between computer based systems
Implementation and customisation of information systems
Monitoring, control and costing
Information flow from customers and sales
Information identification, location and organisation
Implementation and operation of quality systems
Numbering and traceability of machines, assemblies and parts
Information availability and accessibility
Information systems strategy and planning

Table 3.2: A list of nine core issues to information management. Taken from [2].

3.2.3 Dissemination

The dissemination of knowledge is one of the most challenging parts of a knowledge management strategy and as such a number of approaches and knowledge management Systems have been developed to bridge the gap between the need for knowledge and the knowledge source [49]. A typical knowledge management system will have some form of interface that will allow the user to interrogate stored knowledge. Even with these dissemination systems, reports from literature stating that the preferred method of finding information is still asking a colleague [56][57], indicating that there is scope to improve current systems.

Search is one example of this. It is widespread in both organisational knowledge management (intranet/enterprise search) and personal life (Internet search engines) and one could argue that the Internet search providers are highly successful. Enterprise search however is not as successful. The authors of [10][58] and [9] discuss the challenges surrounding this problem. One reason listed is that within an organisation the 'correct' search result is the one 'right' answer, whereas on the Internet the 'correct' result is actually a collection of possible answers that the user can then browse. Other possible reasons include queries having fewer answers, often the user is searching a specific known document, the metric for ranking is different (Google's PageRank [59] versus order by date for example), content structures/technologies are different (cross linking HTML web pages versus CAD or a Microsoft Office Word document), content creators have different mindsets (popularity on the internet versus quarterly finance reports for example), and query formulation (domain specific synonyms or part numbers).

The Internet and Intranet approach to search are different and Intranet technology is yet to see the achievements of the Internet. In the paper [23] based on **Chapter 2: Industrial Context and Scoping Study** of this thesis, this idea is explored further. A linguistic analysis of Airbus Business Search queries revealed that Airbus personnel were more likely to search for ‘things’ when compared to Internet search users. When the Business Search queries where then grouped in terms of their context, the product itself was found to be one of the most popular types of search. The paper concluded that the domain specific nature of search within organisations can provide structure/context to Intranet search, something that is simply not as easily available to Internet search engines. The product structure itself being one means of adding structure/context.

Further examples of augmenting search queries using structure include the use of context, ontologies, and taxonomies [41][60][61]. Taxonomies capture words in a hierarchy of concepts (a golden retriever is a dog which belongs to the canine family of mammals). Ontologies capture relationships between concepts (a golden retriever eats dog food and walks in a park). Capturing and using these relationships provides a wider context to the search that can improve results returned. Working with Airbus, Xie [62], constructed a search engine that leveraged context to provide some structure to unstructured free-text reports. The system, Daedalus, is still in use at Airbus today. Other examples of ontology-based search include Li et al. [63] - in this work, a SPARQL (database semantic query language) approach to design rationale retrieval that uses semantic models and template queries to improve retrieval, again adding context to improve search.

Turning dissemination on its head, push-based strategies aim to predict the information need and send information to the engineer rather than have the engineer perform a search themselves. Campbell et al. [64] discuss the practical and ethical questions surrounding such an approach. There are several problems, one is that predicting a user’s information needs as they perform their day-to-day activities largely relies on closely monitoring those activities. Another problem is the idea of ‘spamming’ engineers with information that they might not appreciate.

Improving the dissemination of knowledge is the primary focus of this thesis and as such this section is one of the most relevant. The key aspect taken from this review is how context can be leveraged to improve search. Literature in this area is, however, focused on expanding search query terms using context. This literature review now moves on to explore model-based definition, a visual means of capturing, storing and disseminating information within the product structure.

3.2.4 Model-Based Definition

The work on the capture, storage, and dissemination of knowledge presented so far is largely associated with the traditional databases and search engine user interfaces.

Model-based definition[11], digital product definition [65], and extended CAD [14] are all examples of attaching annotations to the digital geometric CAD data such that information/knowledge, design rationale for example, can be captured within the CAD package. Effectively, the CAD model itself acts as the means of knowledge capture, storage, and dissemination. The approach is aimed at reducing lead-time, improving quality, lowering cost, as well as improving life-cycle management, and information/knowledge dissemination. Although, there are no quantified examples of these in literature.

model-based definition is a step in the direction of utilising the product structure to provide context to information/knowledge dissemination. It is difficult to imagine a more appropriate location to locate design rational other than directly to the product feature described. The information/knowledge attached are still limited to annotations rather than full documents and the approach is still dependent on CAD software for visualisation [66].

Essentially model-based definition acts as an add-on to the CAD package that enables the management of small snippets of knowledge. Data, information and knowledge are however generated throughout the product life-cycle by both engineers and all the other functions that make up an engineering organisation (Finance, HR, Sales, etc). The re-use of these are just as important to the organisation as design rationale is to the designer. There is then the question of whether the model-based approach is beneficial to the wider knowledge management strategy - including document search. Can the product-related context that has been show to improve document search through query expansion, be used to improve how engineers find documents through a visual three-dimensional user interface that represents the product?

3.2.5 Summary

So far this review has focused on knowledge management literature, the remainder of this section now goes on to explore 'real-world' knowledge management in action within the engineering sectors of aerospace, defence and construction.

3.3 Knowledge Management in Practice

There are often gaps between research and the 'real-world' practices and as such, this section examines knowledge management in practice within three engineering fields. Aerospace (Airbus), defence (Babcock International) and construction (the UK government). Airbus see knowledge management as a key to their success and so invest in internal activities to promote and maximise the re-use of knowledge. Babcock sell knowledge management as a product under their Marine Information and Intelligence arm. The UK Government are in this case a customer of the construction industry and, with the aim of reducing the life-cycle costs of public own buildings, has introduced legislation and working groups that require all new builds to meet building information modelling standards. Standards that aim to product a digital twin of a building and its information. This section now discusses each if these in turn to present a picture of how knowledge management is actually taking place with the aim of provide further foundations for the contributed being presented in this thesis.

3.3.1 Aerospace - The Airbus Group

The Airbus Group knowledge management is the responsibility of a team of individual spread across Europe. The role of the team and wheel is to manage, facilitate and promote knowledge management throughout the organisation. Their activities are centred around the 'Knowledge Wheel' (see Figure 3.2) and an individual member of the team takes responsibility for a segment of the wheel [67]. The segments themselves correspond to a particular knowledge management function, system or service.

Knowledge Management Overall Diagnosis (KMOD) is a service that introduces or improves knowledge management within business functions. The approach is structured and systematic and involves analysing a function knowledge management needs prior to creating a plan to deliver a strategy and action plan to meet those needs.

Effective Knowledge Transfer (ExTra) is initiated on a member of staff's departure. It facilitates a number of meetings between the individual leaving and their replacement and aims to transfer knowledge between the two individuals. The Airbus knowledge management literature states that there typically 5-15 meetings will take place between the two individuals over a space of 6 weeks.

Connect@Airbus is a social networking tool for Airbus staff that spans the entire organisation. It effectively acts like a modern version of a Yellow Pages, connecting people across the group through people and document search. The purpose of these types of systems is to satisfy an individual's information need through connecting them directly with a domain expert.

Professional Networks aims to remove the barriers to collaboration and bring together people with similar expertise. Focusing on the creation or improving of an individual's professional networks, the goal of this practice is to stimulate sharing, learning and innovation across company boundaries. Effectively the practice aims to generate and leverage best practices across company functions through bringing together groups of people with similar interests.

Innovation Management is a service aimed at improving the innovation process from idea creation through to the project proposal. The service takes the form of consultations, workshops and training focused on all aspects of innovation from creative problem solving through to innovation management.

Reuse, Improve and Share Experience (RISE) is the Airbus lessons learned catalogue. The Knowledge Management team provide workshops, training, and services to facilitate best practices in the capture of lessons learned. The RISE system facilitates the capture, search and re-use of lessons learned with the capability of integrating the work-flow into existing processes.

Knowledge Capture and Publishing (KCP) is a methodology designed to facilitate the writing, publishing and maintenance of 'Knowledge Books'. The methodology cycles through the stages of plan, capture, promote, and update and the output is a physical book that are essentially knowledge reference manuals that allow user to find specific bits of information.

Business Search is one of the primary search engines available across the organisation and is available to all Airbus staff with intranet access. Between January and June 2014, the system had over 68,000 unique users who ran over 330,000 unique queries and over 1,000,000 total queries. The system itself uses 16 different data sources, the index consists of 5 million documents, and is 107 GB in size [68]. The system provides search across a number of data sources and implements features such as native permissions (so people only see the documents they have permission to see) and graphical filtering and fine-tuning of results.

Figure 3.2 shows the portfolio of knowledge management activities at the time of writing. The Airbus literature boast that Airbus knowledge management is recognised as "*...an international benchmark improving people and business performance*" and this is not achieved without innovation and continuous improvement. As such, the Knowledge

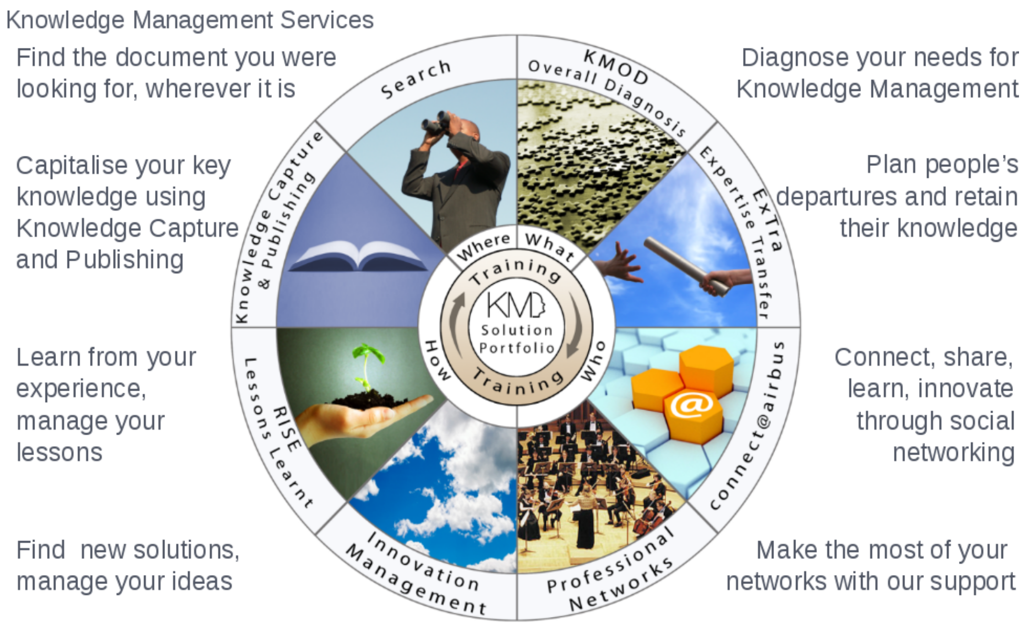


Figure 3.2: The Airbus Knowledge Management Wheel

Management team funds research (as in this thesis) as well as testing and implementing the latest in technologies. Connect@Airbus was one such recent innovation, replacing a traditional Yellow Pages application with a modern social media approach to connecting people. As the organisation changes and as new science and technologies emerge, the portfolio is likely to adapt and evolve however the purpose of the Knowledge Management team remains the same: encouraging the creation, sharing and use knowledge for the benefit of the individual and the organisation as a whole.

The Airbus model of knowledge management shows how tailor made knowledge management Systems provide a means for the effective management of an organisation's knowledge. The fact that the Knowledge Management team are active participants in research also shows that there are still improvements needed in the field.

3.3.2 Defence - Babcock International

Babcock International is a UK based engineering organisation that has been providing bespoke engineering support for over a century[3]. Their marine arm works with naval defences from across the world, including the UK, Canada, Australia and New Zealand[4]. Under Babcock's marine arm, the company lists Information and Intelligence as one of

[3] https://www.babcockinternational.com/Who-We-Are

[4] https://www.babcockinternational.com/What-We-Do/Marine/Asset-Management/Defence-Asset-Management

their products[5].

Babcock break down the Information and knowledge management product into Complex Asset Performance Analytics[6], Collaborative Working Environment[7], Connected Facility[8], Dynamic collaboration for mission critical environments[9], and Digital support for UK Defence[10].

Complex Asset Performance Analytics aims to provide through-life support of complex assets. The complex assets in this case being Type 23 Warship class. Through data models and visualisations of product related data sources (legacy, emergent and future data sets) with the aim of providing a decision support system [69] that improves the 'certainty of delivery' of warship programmes.

Collaborative Working Environment improves collaboration and communication between stakeholders and has been in use by the UK MOD during the overhaul and refit submarines. The service has been expanded over the last 20 years with the aims of reducing costs, improving visibility and exploiting data sets. Babcock boasts the service has reduced the Vanguard refit time by 70-80% and saving the project over £10 million with wider savings resulting from 'business transformations' and cost savings reported at over £28 million.

Connected Facility is a test site for digitally enabled asset management that enables real-time condition monitoring for machine failure prediction and preventative maintenance. The facility also acts as a 'sandpit' for the testing of and 'de-risking' of technology. The aim of the site is to deliver better management of both Babcock and their customers assets, through achieving greater efficiency and more effecting maintenance.

Dynamic collaboration for mission critical environments is a secure SharpCloud for Enterprise[11] implementation for the visualisation of elements such as projects,

[5] `https://www.babcockinternational.com/What-We-Do/Marine`

[6] `https://www.babcockinternational.com/Case-Studies/What-We-Do/Marine/Information-and-Intelligence/CAPA-T23`

[7] `https://www.babcockinternational.com/Case-Studies/What-We-Do/Marine/Information-and-Intelligence/Collaborative-Working-Environment`

[8] `https://www.babcockinternational.com/Case-Studies/What-We-Do/Marine/Information-and-Intelligence/Connected-Facility`

[9] `https://www.babcockinternational.com/Case-Studies/What-We-Do/Marine/Information-and-Intelligence/SharpCloud`

[10] `https://www.babcockinternational.com/Case-Studies/What-We-Do/Marine/Information-and-Intelligence/Smart-Maintainer`

[11] `http://www.sharpcloud.com/blog/sharpcloud-for-enterprise`

risks, data and digital initiatives. The software provides digital visualisations in a collaborative environments with the aim of improving team efficiency and hence the overall project.

Digital support for UK Defence provided a proof of concept in regards to the use of augmented reality within the project life-cycle, with particular focus on the maintenance phase. The system supports maintenance teams through the access to data sets, enhancing skill sets, safety and auditing.

Examining each of these products at the level described on the Babcock website it is possible to determine their underlying information and knowledge management strategy is heavily focused on the access to information and communication between stakeholders. While it may not be the best source of information, the descriptions particularly focus on supporting the decision-making process through access to data. This stands to reason given hierarchical structures and decision making processes associated with defence organisations.

3.3.3 Construction - the UK Government

In 2011 the UK Government Cabinet Office examined the public sector construction of buildings and determined the sector was not returning full value for the money[12]. In response the bBuilding information modelling (BIM) Industry Task Group (2011) followed by the Centre for Digital Built Britain[13] (2015) were formed to drive Building Information Management throughout the public sector[14]. In March 2011, the BIM Industry Task Group released a paper outlining a strategy for the widespread adoption of building information modelling across new build public owned buildings with the aims of bringing supply chains into the process quicker and in a structured and collaborative digital environment. Building information modelling is discussed in greater detail in section 3.4.2 of this chapter, however this section now discusses the UK Government's journey and justification for its roll-out across public buildings in the UK.

In March 2011 the group released a paper[15] that outlined the strategy and includes a description of the first three BIM Levels (see Table 3.3). These levels serve as a good insight into the information and data requirements of both the levels of BIM and the industry as a whole. The levels range from 0 where engineering drawings are unman-

[12] `TheUKGovernmentCabinetOfficeid`
[13] `https://www.cdbb.cam.ac.uk/`
[14] `https://www.cdbb.cam.ac.uk/AboutDBB/Timeline`
[15] `https://www.cdbb.cam.ac.uk/Resources/ResoucePublications/BISBIMstrategyReport.pdf`

BIM Level	Description
0	Unmanaged two-dimensional CAD shared via paper/electronic paper
1	Managed two/three-dimensional CAD adhering to BS1192:2007 and within a Common Data Environment that allows collaboration.
2	Level 1 within a three-dimensional virtual environment with attached data. Representations for Architectural, Structural, Facilities, Building Sources and Bridges.
3	Level 2 plus interoperable data.

Table 3.3: BIM Levels as described by the BIM Industry Working Group

aged and distribution is paper based, the level 3 where engineering drawings, three-dimensional models and interoperable data are all brought together in a collaborative digital environment.

The report also called for the adoption of the data format COBie [70] to be used to hold building asset information as well as act as an index for the two/three-dimensional drawings. The open, non-proprietary format was intended to be act as a 'common medium' that all stakeholders could interact with regardless of their size and IT capability. The format itself consists of spreadsheet data that captures the time stamped spacial location of building assets across the life-cycle of the building. The spreadsheet nature of the format means is accessible without the need for a CAD package and the associated high end computer hardware and the openness results in suppliers not having to all adopt the same proprietary software package.

The life-cycles of buildings and their assets are decades if not hundreds of years long and at these time scales, details such as file formats and software and IT hardware life-cycles become a challenge in their own right. While the COBie format and its open, non-proprietary format is a step in the right direction, computer science research is yet able to address the challenges of long life-cycle software.

The key point to note with the UK Government's call for BIM is the underlying strategy of integrating information with a three-dimensional representation of the product structure. Akin to model-based definition, the means of capturing, storing, and disseminating information is the product itself.

3.4 Engineering Information Management

Engineering information management systems are knowledge management systems tailored to specific engineering applications with product life-cycle management and product data management being the main two examples [50][71]. Product life-cycle management aims to aid the management of the product as it moves through the phases of its life while product data management occurs within the product life-cycle management system to capture and manage product data created during the various life-cycle phases. Also relevant to this thesis, are the fields of model-based definition and building information modelling. Model-based definition is a technique that allows engineers to annotates CAD models such that elements such as design rationale are captured within the CAD model itself. Building information modelling systems have emerged from the construction industry and managing the digital life of a building alongside the physical life-cycle.

3.4.1 Product Life-Cycle Management and Product Data Management

One view of the product's life-cycle is that it moves through five phases, imagination, definition, realisation, use/support, and disposal [72]. Within each of these phases there are differing information needs, ranging from standards through to supplier contact details [39]. It is the product life-cycle management systems role to manage the transitions between the phases and the product data management systems role to manage the data, information and knowledge generated.

One of the key experts in the field of product life-cycle management, John Stark writes about the PLM Grid [72], a five by ten grid with the five life-cycle phases running against the ten components addressed by product life-cycle management.

Objectives and metrics are the high level business objectives and the metrics against which they are measured. Examples of metrics include time, cost, quality, and quantity for examples. These objectives are interdependent and often a primary focus on one will impact another - quality dropping as a result of increasing quantity for example.

Management and organisation relate to the structure of the organisation and the management of resources. Examples of different types of organisations include functional, geographical, hierarchical, project, product-focused, and virtual. Resources include people, equipment, product, product data, and facilities for examples.

People perform a range of activities within a product's life. These include business analysis, programming, management, service engineers, and executives. People are often organised into different departments, therefore aspects such as language, culture, information needs, and location are all likely to vary.

Methods are techniques employed to aid specific parts of the life-cyle. Concurrent Engineering, Six Sigma, Business Process Management, Continuous Improvement, and knowledge management to name a few. The purpose of these methods is to improve and optimise the organisation such that it meets the demands of the market in which the organisation operates.

Facilities and equipment are the machinery and buildings required to design, manufacture, store and sell the product. Understanding the equipment available, the facilities to hand and the logistics between them is vital. An example of this from Airbus is how the A380 wings are manufactured at Broughton in the UK before being shipped to Toulouse in France for final assembly. The distance may not appear sensible until you factor in lead times in the final A380 assembly process.

Product life-cycle management applications exist in various forms with various actual purposes and specialities. Computer Aided Design (CAD), Bill-Of-Materials (BOM), parts catalogues, 3D scanning and printing, databases, and finite element analysis being some examples. Communication or easy transfer of data between these applications would be ideal however the reality, as shown by Hicks et al. [2], is that it is not always the case.

Product data management systems are a specific component of the product life-cycle system that specifically manages product data. Stark describes eight components of a data management system: information warehouse, information warehouse manager, infrastructure, system administration manager, interface module, product and work-flow structure definition module, work-flow control module, and information management module.

Product data *"defines and describes the product"* and are created throughout the product's life-cycle. Examples include: CAD geometry, process plans, standards, customer requirements, software, change data, and costings. Management of data introduces rules around access, granularity, security, standards, and lifetimes for example.

Processes are the activities that take place within the life-cycle. High level examples include design, manufacturing and sales. Detailed examples include parts purchasing, testing and assembly, prototyping, packing, machine set-up, inspection, and project management. Processes are interconnected with some dependent on others finishing first.

Products are effectively the artefact being managed and, in most cases, sold with the aim of making a profit. Important factors regarding products include number and range of products, part numbers for identification and traceability, and the product structure, bill-of-materials and assembly process.

According to Stark, product life-cycle management must manage and cater for each of these if it is to be successful. This is a complex range of components that cover all aspects of the organisation and not just the product itself.

Examples of software packages include Siemens Teamcenter[16], Autodesk Fusion LifeCycle[17], and PTC Windchill[18]. These three examples are said to be able to manage, facilitate or integrate: bill-of-materials, configuration and change, requirements, quality, platform structures, product data, collaboration, integration & standards, supplier, simulation, manufacturing process, service life-cycle, environmental compliance, and costing. In essence, product life-cycle management systems aim to be a one stop shop for all the product information needs.

These systems are however, still largely based on databases, share folders structures and text-based search engines for the finding of documents. As such, improvement can still be made in their performances in terms of knowledge dissemination.

3.4.2 Building Information Modelling

The UK Government's drive for building information modelling across public buildings was covered in section 3.3.3 of this chapter. This section explores building information modelling in terms of the fundamentals behind such systems. These systems generate and maintain a digital representation of a building the lives alongside the physical building. The digital model is maintained throughout the life-cycle of the building and act as a record of the physical structure of the building along with its assets and maintenance/repair history. The model can then be interrogated for information throughout the life-cycle with the aim of improving collaboration between stakeholders and reducing the life-cycle cost of the building.

In the BIM Handbook [15], the authors provide a detailed overview of a range aspects of building information modelling and the various stakeholders involved. Using

[16] https://www.plm.automation.siemens.com/en/products/teamcenter/
[17] http://www.autodeskfusionlifecycle.com/en/
[18] https://www.ptc.com/en/products/plm/plm-products/windchill

the handbook as a guide, the section will now discuss the fundamentals of building information modelling, starting with the technology before briefly exploring the range of stakeholders and their interactions with the system.

Building information modelling technology can be thought of an intelligent and integrated CAD. CAD models are constructed using geometric data however, these 'objects' also embody functional and rule data alongside the geometric data. The result is a system that can update and monitor itself such that, walls will re-size themselves to butt to roofs and doors will always perfectly fit into walls. In addition to this, objects are able to check themselves against predefined rules and flag breaches. If the design changes that manufacture is no longer feasible for example.

Objects are also able to communicate both with each other and external applications, this means two things. Firstly, changes in objects automatically propagate through the model resulting in a consistent model. Hierarchically grouping components supports this dissemination of attribute data, if the weight of a wall changes then this automatically updates the weight of the building for example.

The second form of communication allows for external applications to load, view and interrogate the model resulting in a system that dynamically adapts to stakeholder needs. This leads on to how different stakeholders interact with the model, as the system is able to adapt to stakeholder requirements. Environmental engineers accessing and exporting a building's energy efficiency for example. To achieve this, however, the means of communication need be set such that different applications are able to interpret data. The BIM Handbook suggests either using one vendor for the entire building information modelling system or sticking to open standards such that multiple vendors can be used.

Building information modelling aims and objectives are envisioned throughout the building life-cycle from concept through to service. During concept, feasibility and design the virtual model enables reductions in costs and times through functional analysis of designs earlier in the process and prior to detail design. In fact, modelling allows earlier collaboration of design disciplines, defect checking, and design visualisations. Other benefits include automatic low-level error corrections, cost estimation and two-dimensional drawing generation at any stage of the design process.

During construction these systems are aimed at simulating the construction process itself, allowing engineers to move through a simulated construction. This results in early detection of errors and quicker reactions to on-site problems. The design model can also be used to directly fabricate and manufacture components and synchronise procurement.

Finally, during the actual use of the building, these systems enable facility and asset management. Examples of this include simply visualising where electricity cables run behind walls and interrogating the maintenance history a particular asset.

Each of these generates benefits through providing access and collaboration throughout the building life-cycle. The building information modelling system acts as a comprehensive digital duplicate of the physical building that captures the building's data, information and knowledge and provides access to it to those who need it, in the most intuitive context, the building itself.

AutoDesk Revit[19] is an example of building information modelling software on the market. The package claims to manage the building life-cycle from concept through, planning, design, construction, and in-service while boasting a host of features including parametric modelling, work scheduling, three-dimensional visualisation, annotation, laser scan point cloud connecting, performance analysis, structuring modelling, and cloud rendering. Revit aims to be a one-stop-shop for the digital management of a building's life cycle, however, whether the package is able to continue to support a building over its full life-cycle of a build is a long way from being tested.

3.5 Discussion and Conclusion

This section explored the literature and 'real-world' of knowledge management from an engineering perspective. The knowledge management landscape presented is one that identifies and prioritises the need for the capture, storage and dissemination of organisational data, information and knowledge while acknowledging the need for innovation and further improvements.

While knowledge management as a whole is concerned with knowledge capture, storage, and dissemination, the remit of this thesis is such that the focus here is to improve the dissemination of knowledge. Search engines are one of the key means for dissemination information however and there are several examples showing how expanding search queries with context can improve search. There are also examples of techniques for managing specific pieces of information using the product structure itself, whether that is model-based definition or building information modelling. These are three-dimensional user interfaces being used for information dissemination. There is

[19] `https://www.autodesk.com/products/revit`

however a gap in knowledge in terms of document navigation via the product structure, i.e. can re-representing a collection of documents via visual representation of the product structure or model-based information navigation, improve engineering information access and knowledge discovery?

The key findings from the literature review are then:

- The dissemination of knowledge is largely based on Internet technology and found lacking
- The use of context to expand search queries can improve search
- A group of systems has begun to provide three-dimensional user interfaces that leverage the product structure as a means to capturing specific pieces of information
- A gap in knowledge exists in terms of the navigation of document collections via the product structure: model-based information navigation

Prior to finishing this chapter, it is worth noting that Chapters 5, 6, and 7 all contain further reviews of literature that are more specific and detailed to the Research Questions 1, 2, and 3 respectfully. The next chapter begins this journey by describing the methodology employed to investigate this model-based approach. In this next chapter, the work output from **Chapter 2: Industrial Context and Scoping Study** and the literature review presented here are brought together to form the research questions that this thesis sets out to answer.

Chapter 4

Aim, Methodology and Research Questions

4.1 Introduction

The key output from **Chapter 2: Industrial Context and Scoping Study** showed how Airbus personnel search for product related information. **Chapter 3: Literature Review** then showed how enterprise search engines, used for the dissemination of knowledge and information, are being improved through the expansion of search queries using context. Chapter 3 also shows how model-based definition and building information modelling systems are placing a visual representation of the product at the heart of accessing specific pieces of information. The review of the Airbus knowledge management strategy also showed how specific tailor made systems can be used to satisfy specific knowledge management needs.

Bringing all this together, this thesis proposes knowledge-based information navigation as a strategy for improving information access and knowledge discovery: a means of navigating document collections via a three-dimensional representation of the product structure. The specific aim of this thesis then is: *The development of a framework for model-based information navigation and knowledge discovery in large engineering organisations.*

This chapter begins by outlining the research methodology used to move from the scoping study and literature review through the formulation of research questions that ultimately allow for the aim of this thesis to me met. Once the methodology is explained, this chapter briefly discusses the technology platform used throughout the thesis and how

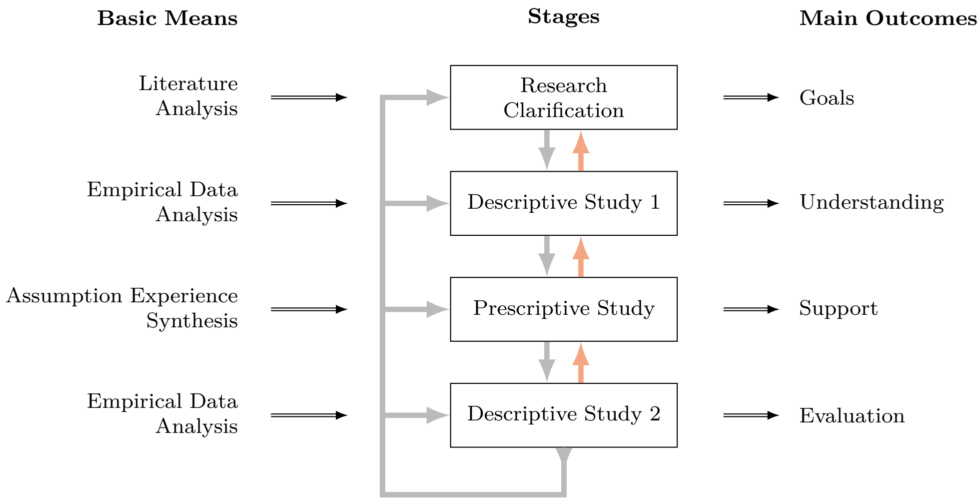


Figure 4.1: The Design Research Methodology Framework [1]

its development helped form the research questions. Then the research questions themselves are described before the second half of the chapter identifies evaluation techniques and outlines how each of the research questions will be answered through literature review, study and evaluation.

4.2 Design Research Methodology

The design research methodology [1], is a framework for design research aimed at providing rigour to research activities and it is the framework used throughout this thesis. Figure 4.1 shows the framework and its four stages. During Research Clarification, the overall aim/research goal is determined. During Descriptive Study 1, activities are undertaken to better understand the landscape within which the research aim sits and to clearly identify the areas that will need addressing, typically a literature review and ending with the list of research questions to be addressed. During Prescriptive Study and for research associated with the design, development and evaluation of a software tool, this is where the software is developed. During Descriptive Study 2, the software tool is evaluated against the original research aim/research goals.

Figure 4.2 is then how the design research methodology is applied in this thesis. The aim of the thesis was determined during research clarification through a combination of literature review, review of the original research proposal and through clarification with stakeholders within the Airbus Knowledge Management team. Descriptive Study 1 involved a further detailed literature review (**Chapter 3: Literature Review**), the scoping study (**Chapter 2: Industrial Context and Scoping Study**) and an exploration of technology platforms (outlined in Section 3 of this chapter). At this time, there was a continuous iteration between Descriptive Study 1 and Research Clarification as the thesis aim and research questions evolved as a result findings from literature review, scoping study, technology exploration, and stakeholder feedback. The output from Descriptive Study 1 was a set of research questions addressed in this thesis and presented in Section 4 of this chapter. The first three research questions related to how one would build the system and as such, formed the Prescriptive Study stage. Research Questions 1, 2, and 3 are answered using a combination of literature review and study, presented in chapters 5, 6, and 7 respectfully. Those three questions then allow for the final software platform to be developed and tested in Descriptive Study 2. Chapter 8 describes the designing of the final study, chapter 9 presented the results and chapter 10 reviews the thesis in its entirety.

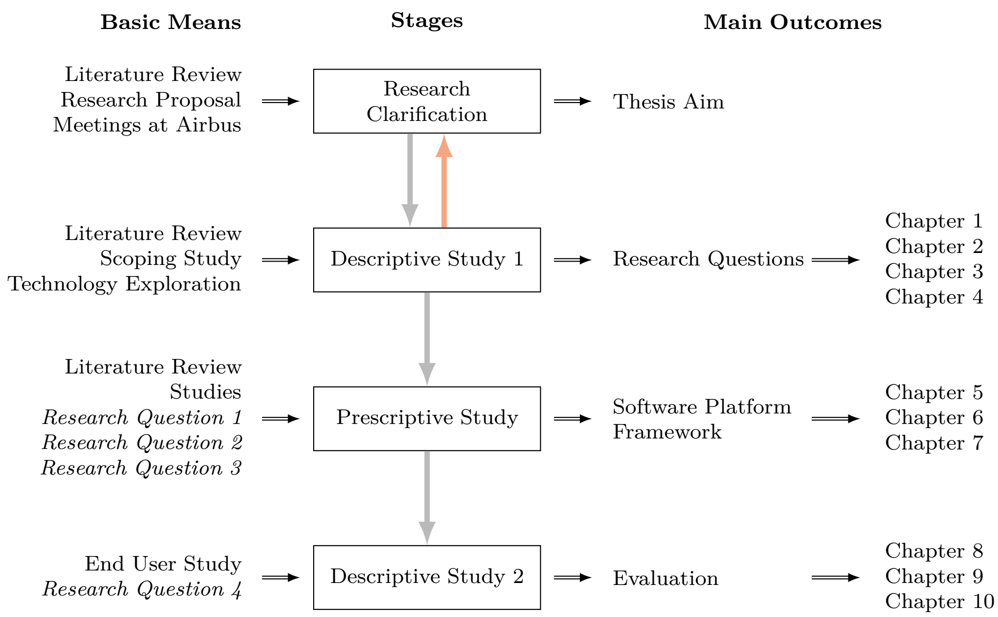


Figure 4.2: The Design Research Methodology Framework as applied in the thesis

As indicated in the research methodology, forming the research questions involved the scoping study (chapter 2), literature review (chapter 3), and an exploration of technology. The next section in this chapter now briefly covers this technology exploration such that the research questions can be presented.

4.3 Technology Platform

At this point in the thesis, the technology platform used throughout the thesis is discussed. The development of the prototype system was iterative and took place over the length of the thesis. The cyclical nature of the iterative development means there is not a natural point to introduce the technology platform within the linear narrative of the thesis. However, along with the literature review and scoping study, the exploration of technology supported the development of the research questions and as such the technology is covered here, prior to the research questions being introduced. Some elements of the details presented in this section are then presented in detail in later chapters.

The size and complexity of Airbus products means that high-end CAD packages and hardware are required open and view CAD models and as such, only a few Airbus engineers have the necessary technology and training to interact with full three-dimensional representations of their products. This is something that many at Airbus would like to change and Airbus are investing in tools to generate light-weight CAD representations that are accessible to more standard software/hardware, democratising the CAD model. From a knowledge management perspective, it also makes sense to design systems that benefit the maximum number of people. As a result, Internet technology was selected as the platform for visualisation. The main justification being:

- Most modern computers have web-browsers
- Web-browsers off an easy solution to cross-platform compatibility
- Internet technologies are tailored to this type of application
- The advent of WebGL has allowed for three-dimensional acceleration within the web-browser

The third and fourth point require further explanation and this is now provided.

Internet Technologies touch most of our lives on a daily basis and are implemented in a number of applications that range from simple landing pages that display basic text and images, to shopping, news, games to name a few. As a result, the technologies behind these applications are robust, dynamic and tailored for the mass access of information, products and services. The rapid growth of Internet innovation is frequently attributed to the Open Source nature of the underlying technology.

Linux or Unix-based operating systems run on the vast majority of the worlds computers, from mobile devices through Android and IOS through various flavours of Linux server distributions. Linux servers are tailored for multi-users and given their Open Source licensing, are highly customisable. The two most popular web servers used are Apache and NGINX, there are many debates over which is better but for the purposes of this thesis Apache was chosen as the author has more experience in developing with it. Pairing up with Linux and Apache, MySQL and PHP are a classic set of technologies that form a LAMP server. Linux manages the hardware, Apache manages the internet, serving up web-pages, MySQL is a database for the structure and storage of files, and PHP is the interface between the server side technologies (database and the operating system file system) and the web side of the set up. All these are freely available, Open Source, robust and dynamic and should provide a dependable platform for development. This back-end architecture only works if it can deliver a three-dimensional graphical front end.

WebGL [16] is a relatively new technology that used JavaScript to re-write the Open Source graphics library OpenGL [73] and provided the web browser with access to the graphic hardware on client side computers. This essentially gives the web browser three-dimensional graphics acceleration. WebGL is quickly being adopted and examples of its use include: games[1]; interactive visualisation[2]; and even CAD packages[3]. Again the technology is Open Source and freely available. This technology then offers a means by which democratisation of the CAD model can be achieved.

The literature review revealed that unless one is to use a single vendor for all knowledge/product life-cycle management/building information modelling needs, that the only route is to use one where the intermediary file types and systems are meet Open standards. This technology architecture achieves this, therefore, the system development will take place using highly customisable, robust, and freely available Open Source technologies.

The development of a model-based information navigation System took place iteratively over the entire length of the project. The iterative approach allowed for the system to be refined and 'evolved' toward a suitable solution. The iterations involved the design, build and informal testing/end user feedback of a number of technologies and interfaces. The testing/feedback was provided by academic and Airbus engineers. The first three research question were derived from this process as the system evolved,

[1]www.webgl.nu. Last visited 2017-09-18
[2]www.threejs.org. Last visited 2017-09-18
[3]www.webgl-publisher.com. Last Visited 2017-09-18

challenges arose that warranted research and study to answer. These then became the research questions presented later in this chapter.

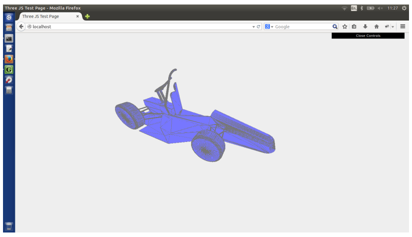
Figures 4.3a, 4.3a, 4.3b, 4.3c, 4.3d, and 4.3e were presented in [74], and show the first prototype that was used to determine the feasibility of the model-based approach. In this early version, documents are represented as point-of-interests (bright green cubes) situated around a racing car. As the user moved around the car, the results showing the results panel (Figure 4.3c), filter based on the point-of-interest visible in the model panel. The findings from the paper showed that the point-of-interest (considered state-of-the art), were not easy to navigate and introduces the possibility of the component-of-interest (Figures 4.3d and 4.3e) - an approach where documents are connected directly to components themselves.

Figures 4.3f and 4.3g show the evolution from a basic visualisation to a more complex detailed visualisation that also introduces the functionality of clicking on components to access a list of associated documents. Figures 4.3h and 4.3i show the application of visualisation colour theory to the visualisation and Figure 4.3j shows the final version of the system, with the introduction of user interface design theory, hosting means to navigate the model environment (the green and red panels).

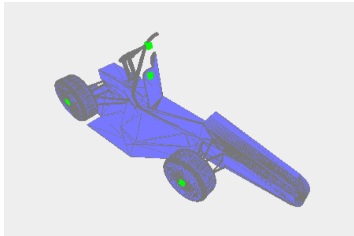
During this exploration of technology, a number of challenges arose:

- how do documents relate to the product structure?
- how navigate within the model-based environment?
- how to display information within the model-environment?

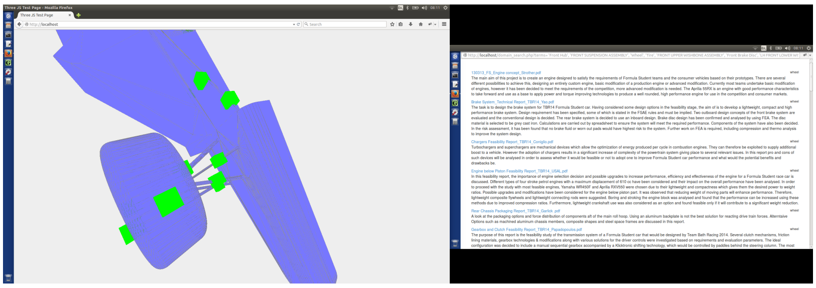
The first derived from the finding that situating the points-of-interest within the model-environment was non-trivial task. Point-of-interests are situated with a single x, y, and z co-ordinate, however, this single point in spaces does not reflect the true relationship between the document and the model, a document might relate to a wheel but rarely a single point. The second challenge related more to the user moving around the model environment. If the document is situated deep in the heart of the product structure, how does the user find it and interact with it? The third and final challenge again relates the point-of-interest. If information exists within the model environment, the user needs to be made aware of it, i.e. they need to see it exists to be able to find it. This third challenge then relates to what documents should look like such that they are easily seen. These three challenges formed the basis of the research questions now presented.



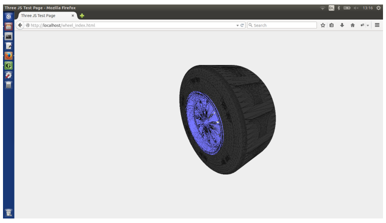
(a) First model-based information navigation prototype



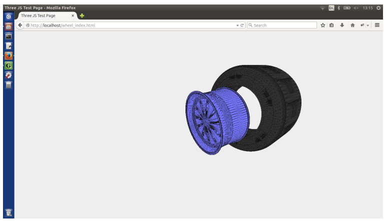
(b) An early model-based information navigation prototype



(c) An early model-based information navigation prototype including results panel

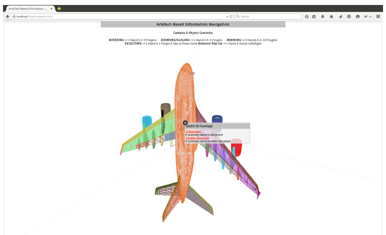


(d) An early model-based information navigation prototype demonstrating the component-of-interest concept

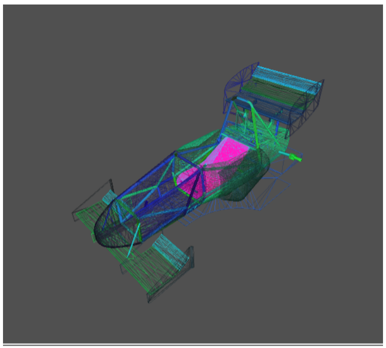


(e) An early model-based information navigation prototype demonstrating the component-of-interest concept

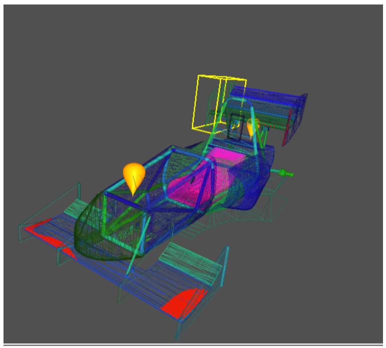
(f) An early model-based information navigation prototype



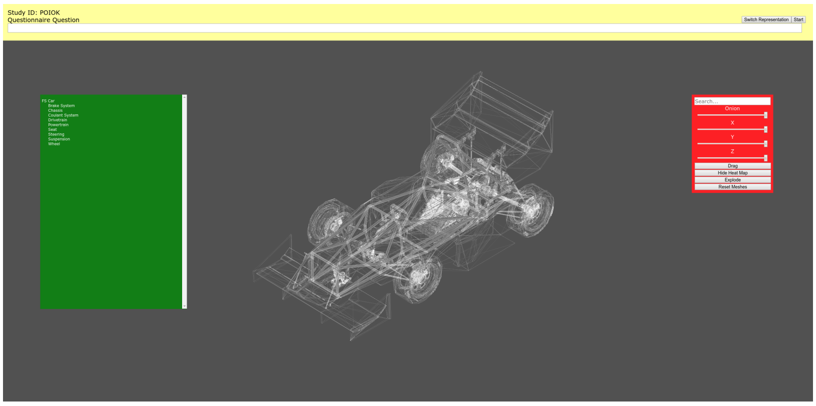
(g) An early model-based information navigation prototype with click to open reports implemented



(h) A later version of a model-based information navigation prototype



(i) A later version of a model-based information navigation prototype with visual information objects implemented



(j) The final model-based information navigation prototype

Figure 4.3

4.4 Research Questions

Evaluating the concept of model-based information navigation is divided into two components, the first is designing and building the system such that it operated in a manner that does not inadvertently affect the second component, measuring the affordance of such as approach to information navigation. Based on the findings from the iterative development of the prototype systems shown in Figure 4.3, the literature review and scoping study, a number of challenges were discovered. This section of the thesis now formalises these four research questions.

4.4.1 Research Question 1: What are the most appropriate techniques for a model-based approach to document indexing?

The traditional method of accessing data sets is the search engine and indexing, the method already applied by the Airbus Knowledge Management team through the Business Search system. It stands to reason then that the document access side of the system mirror this approach such that the model-based approach can effective plug into a traditional architecture that brings with it, access to a range of data sets. This still however forms the first research question: are the appropriate techniques for a model-based approach to document indexing? The likes of taxonomies and ontologies aim to improve search through providing structure to otherwise unstructured documents and so does the actual product structure achieve any benefits over a traditional search engine? From the user interface perspective, this is akin to where on the three-dimensional model should documents be found?

4.4.2 Research Question 2: What are the most appropriate techniques for navigation information within a model-based virtual environment?

The second research question and aspect to delivering model-based information navigation is how the user navigates the three-dimensional virtual environment that is the virtual product structure. CAD has particular navigation techniques that are aimed at supporting the engineering in interacting with a design environment but these might not be appropriate for an information navigation system. It may be that the navigation techniques of computer games, three-dimensional maps or traditional file system navigation are more appropriate. Research Question 2 then aims to answer: what are the most appropriate techniques for navigation information within a model-based virtual environment? To return to the user interface, how does the user move to find and select documents within the model environment?

4.4.3 Research Question 3: What are the most appropriate techniques for displaying information within the model-based virtual environment?

The third research question explores the development of the system relates to how information is displayed within the three-dimensional virtual environment. There are occasions where documents relate to individual components and other times when components will relate to groups of components related by function or proximity. These occasions lead to the question: what are the most appropriate technique for displaying information within the model-based virtual environment?

4.4.4 Research Question 4: How does model-based information navigation improve engineering information access and knowledge discovery?

The final research question examines the concept of model-based information navigation as a tool for document search within a knowledge management portfolio. Effectively the aim is to discover how a three-dimensional visual representation of the product improves the way engineers access information and the benefits from a knowledge discovery perspective. How does model-based information navigation affect engineering information access and knowledge discovery?

4.4.5 Summary

This section identifies the four research questions that this thesis aims to answer. The chapter now moves on to outline the technology that will be used to develop the system and the evaluation techniques that will be employed to answer the research questions.

4.5 Evaluation Methods

A visual system for the interaction with information sits on the barrier of a number of research fields, document indexing, information retrieval, machine learning, human-computer interface, and information foraging to name a few. It is then paramount that the research methods employed mirror these fields. Fortunately, the work of Catarci et al. [75] covers this type of application. Their paper discusses the evaluation of information retrieval from a human-computer interfaces perspective and provides a thorough overview of suitable evaluation methods. This section is largely based on the work presented by Cararcu et al.

The paper begins by outlining how the traditional measures of precision, recall and f1-measure may not always be appropriate when examining the usability of IR systems. The paper goes on to outline an evaluation framework for usability of IR systems. This framework focuses on the participants, tasks, measures and evaluation methods. This document will discuss this framework in relation to the model-based information navigation system and present a strategy for the evaluation of each of the research questions in turn.

4.5.1 Precision, Recall and F1 Measures

The predominant measures used in IR evaluation are precision and recall. Given the widespread use it is only right to discuss what these measures represent and limitations as a true evaluation of the usability of IR systems. The application of precision, recall and f1-measure require a data set that has documents with predetermined classifications. In this manner if document X is classified as related to the fuselage and document Y to the port aileron then the search query 'fuselage' should return document X but not document Y. This is in essence what the precision and recall measures present.

From [76] the formal definitions of precision and recall are as follows:

$$Precision = \frac{relevantDocuments \cap retrievedDocuments}{retrivedDocuments} \tag{4.1}$$

$$Recall = \frac{relevantDocuments \cap retrievedDocuments}{relevantDocuments} \tag{4.2}$$

Precision is the fraction of the number of true positive results in the total number of results returned. The higher the value, the lower the number of false positive documents returned. Whereas recall is the fraction of correctly returned documents from the complete set of relevant documents. The harmonic mean of these two measures gives the f1-measure or f1-score and this is given as:

$$F1\ Measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4.3}$$

Being the harmonic mean, the f1 measure tends towards the lower value of precision or recall and so the f1-measure will only be high if both precision and recall are high. This is the desired state but is a key challenge in building IR systems.

The problem with precision and recall in usability studies, as outlined in [75] is that to generate these measures the data set must be classified into predetermined classifications. These predetermined classifications are fixed and so cannot take into account additional variables like the user profile, the information needs or the search context. In an enterprise-wide information navigation system the users are likely to range from engineers with engineering related searches to HR personnel with HR related searches. The correct classification of a document can differ in the contexts of these two searches and traditional precision and recall cannot measure this. Catarci et al. discuss precision and recall variations that can account for the user and information and these will be discussed in more detail in the next section.

4.5.2 Human-Computer Interaction Meets Information Retrieval

Catarci et al. [75] explores the world of information retrieval from the perspective of human-computer interaction. There is not a great deal of work in terms of evaluation methods in cross over between these two fields however, the model-based approach sits in the intersection and as such, the work presented by Catarci et al. heavily influences the evaluation methods used throughout this thesis.

Catarci et al. breaks the evaluation of IR system usability into participants, tasks, measures and evaluation and discusses each in detail. This section will now examine each of these and the relevant detail in relation to the evaluation of a model-based approach to information navigation. The outcome of this section should be a deployable strategy for testing the four research questions.

4.5.3 Participants

Catarci et al. discusses three types of participants. HCI experts, users and surrogate users.

HCI Experts: During the design stages it is common for HCI-IR systems to use expert-based usability inspection methods of evaluation (like heuristic evaluation and cognitive walk-through) using HCI experts. Heuristic evaluation is a process of examining User Interfaces (UI) for obvious problems and involves experts examining the systems and summarising their opinions of the positive and negative aspects of the UI

[77]. Cognitive walk-through involves experts walking through set tasks and consider how easy it would be for new users to learn to use the interface [78]. Together these can be fed back into the system design and remove any obvious flaws prior to any users seeing the system. Informal versions of these tests were used throughout the iterative development of the prototype systems shown in Figure 4.3.

Users and Surrogate Users: It goes without saying that end user evaluation is useful in understanding system usability. Studying end users generates an evaluation environment that is the closest match to the real world environment for which the system is intended. It can be difficult to recruit end users for testing and so it is acceptable to use surrogate users in these cases although care should be taken to use surrogates as close to the actual users as possible.

So far in this thesis, the users have been discussed in terms of the Wing In-Service team based at Filton in Bristol, UK. During the project however, the front-line nature of the In-Service team resulted in the team not being available to participate in studies and as a results, surrogate users and data are used throughout the thesis. These are discussed in more detail in the chapters relating to the studies themselves.

Tasks: In designing the tasks for IR usability evaluation it is important that the task be designed to mimic the real world usage of the system. To achieve this at Airbus, the best approach would be to discuss search with the personnel that we are currently in contact with. From these a number of simulated real world scenarios can be formulated and combined to generate the tasks for evaluation. [75] does note however that it is important not to design tasks that "*…are not overly involving or tedious*".

4.5.4 Measures

Catarci et al. discusses a number of measures with regards to evaluating the usability of HCI-IR systems, these range from the standard measures of effectiveness, efficiency and satisfaction to the more user centric measures of interaction, user characteristics, information needs and a few others. This section will discuss each of these in relation to a model-based approach to information navigation.

Standard Measures - Effectiveness: The traditional measures of measuring effectiveness are precision and recall but as discussed earlier, these are not the most appropriate for evaluating the usability of IR systems. Catarci et al. discusses interactive precision, interactive recall and interactive TREC precision which may offer a viable alternative to the traditional measures, however, [75] also suggests eliciting perceptions of performance from user studies and this may be the better option for the evaluation of a model-based approach to information navigation.

Standard Measures - Efficiency: Taken directly from [75], the traditional means/measures for efficiency are below. User perceptions of efficiency can also be used. In terms of evaluating a model-based approach to information navigation all these measures could be combined. Given a set task, the measurements of total time to complete the task, time taken manipulating the model and time sifting results could be automatically measured by the system.

- The overall time the user takes
- The time the user takes doing specific things
- The time the user takes in specific or different modes

Standard Measures - Satisfaction: From [75] "*...satisfaction can be measured by eliciting self reported data from users about their level of contentment, fulfilment or gratification as a result of using or interacting with the information retrieval system* ".

User Interaction: Taken directly from [75] again. The interaction measures listed below cover aspects of using the system. While query length would not be relevant in a model-based approach to information navigation, the others are and the system could be automated to record these for each user and task. [75] also states that interaction measures apply when evaluating effectiveness. See Table 4.1 for a list of the measures and the resultant inference used in this thesis.

User Characteristic: The model-based approach to information navigation is intended for a number of users with different roles within Airbus and as such it is important to understand how the usability might differ between those different users. [75] suggests recording attributes like sex, age, profession, role, experience using similar systems and search experience and to make that specific to Airbus characteristics like job title/function and location could also be included.

Information Needs: The information needs are task related, i.e. the purpose of the search and examining these can allow for the better understanding of the effectiveness

Measure	Inference
Number of queries	A high number of queries indicates the user is unable to satisfy information need. The higher the number of queries, the less effective the system.
Number of search results viewed	The higher number of results viewed, the higher the amount of effort required. Effort should be minimal hence, the more effort required, the less effective the system.
Number of documents viewed	The higher the number of documents viewed the higher the amount of effort required. Effort should be minimal hence, the more effort required, the less effective the system.
Number of documents saved	The higher the number of documents saved indicates a higher number of possible 'correct' results. Reviewing each of these requires effort and as such it indicates a less effective system.
Query length	The longer the length of a query (i.e. the number of words) inputted indicates a greater effort in filtering search results and a less effective system.
Appropriate combinations of the above measures	Number of documents viewed divided by number of documents saved being a measure of the effort involved in finding the right answer, for example. A high ratio of documents opened to documents saved would represent a low effectiveness. E.g. dividing number of documents viewed by the number of documents saved measures the effort involved in finding the right answer. A high ratio of documents opened to documents saved would reflect a low effectiveness.

Table 4.1: Interaction measures for evaluating effectiveness.

and efficiency measurements. Those information need measures related to a model-based approach to information navigation are task-related measures (task-type, task-familiarity, task-difficulty/complexity) and topic-related measures (topic familiarity and domain expertise).

Data Capture [75] identifies system logging, questionnaires, interviews and a process called thinking-aloud as methods of data capture.

System Logging: Generating system logs would involve modifying the system to capture key processes like number of clicks, time between certain actions, etc. There are a number of JavaScript libraries that perform 'logging-off-the-shelf' but for the purposes of evaluating a model-based approach to information navigation it may be necessary to construct a customised logging system to capture the specific requirements of a model-based approach to information navigation and the usability evaluation process.

Questionnaires: It is common for questionnaires to be used during IR usability evaluation and for the a model-based approach to information navigation evaluation these will be the main source for obtaining user perceptions of the system. A questionnaire could be completed at the beginning and end of a given task, at the beginning to collect user characteristics and at the end to measure the system effectiveness and user satisfaction.

One particular questionnaire from literature that is particularly appropriate is IBM's Post-Study System Usability Questionnaire (PSSUQ) for lab-based evaluation and Computer System Usability Questionnaire (CSUQ) for non lab-based evaluation [79]. Researchers at IBM combined several existing evaluation methods to form a standardised questionnaire to be used across the organisation. The questionnaire asks users to provide a numerical answer (between 0 and 7) to a range of specific psychometric questions. The PSSUQ is designed to be given to users once they have been asked to perform a number of set tasks using the system. This aligns well with the qualitative user opinion evaluation of the visual interface.

Interviews: In terms of a model-based approach to information navigation evaluation, interviews may be best used to capture HCI expert feedback during heuristic evaluation and cognitive walk-through exercises.

Thinking Aloud: This is a process where users are asked to vocalise their thoughts while completing set tasks. Catarci et al. notes that it can be difficult for users to actually focus on both the task in hand and the vocalisation of thoughts and so it is not always the best approach to use.

4.5.5 Summary

This section outlined a range of evaluation methods appropriate for the evaluation of a three-dimensional graphical interface to document navigation. While the vast majority of the section is based on the work of Catarci et al., there are relatively few publications in the field and Catarci et al. have published a thorough and complete overview of the field.

There are, however, several components to this thesis (all four research questions) that require evaluation and as such, the next section breaks down each of the research questions, outlining how they will be answered and the specific evaluation method employed.

4.6 Research Plan

This section outlines the research plan employed to answer the four research questions. Answering each of the questions forms a standalone piece of work and as such, the full details of the methods, participants, tasks and measures are covered in corresponding chapters identified at the end of each subsection.

4.6.1 Research Question 1: What are the most appropriate techniques for a model-based approach to document indexing?

Answering this will involve a literature review and study that examine the field of information retrieval, document indexing and document classification. The appropriate techniques being measured against a traditional search engine and as such, it stands to reason that the study compares a model-based approach to document indexing with a more traditional search engine. Traditionally, search engines are evaluated using the measures of *precision*, *recall* and $f1$-*score* [80]. Answering Research Question 1 then involves the research, design, development, and evaluation two search engines using the traditional evaluation methods as a comparison. One with a traditional document indexing approach and one that leverage the model structure. This work is presented in **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**.

4.6.2 Research Question 2: What are the most appropriate techniques for navigation information within a model-based virtual environment?

Answering this question will involve a review of literature and examples of navigation within similar systems. The results from this will be evaluated using an end user study using surrogate engineers and surrogate use-cases/tasks. Measures will be taken from questionnaires and records of user interactions with the system, the system will log user clicks and their timing. The details of all this is covered in **Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments**.

4.6.3 Research Question 3: What are the most appropriate techniques for displaying information within the model-based virtual environment?

Answering this question involves a range of semi-structured interviews with academic engineers and Airbus staff to determine the ways in which information is related to engineering artefacts. Design theory will then be reviewed to understand how best to depict these relationships and a range of visual information object will be created. Finally, the visual information objects will be evaluated using a questionnaire and a range of academic and industrial engineers. The questionnaires will show images of each of the visual information objects and along with the set of information types derived from the structured interviews and from these, the appropriateness of each of the markers will be determined. The details of this is covered in **Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation**.

4.6.4 Research Question 4: How does model-based information navigation improve engineering information access and knowledge discovery?

Answering this research question will involve selecting the appropriate outputs from the first three research questions and using them to develop a final prototype system. The evaluation will be based on that employed for answering Research Question 2. Surrogate users will be given access to the system and given a number of information search tasks to complete. The system will take interactions with the interface (mouse clicks and timing) and at the end of the study the participants will complete a questionnaire based on IBM's Computer System Usability Questionnaire. The details of this is covered in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** and **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**.

4.6.5 Summary

This section outline the plan to answer each of the research questions in turn. With the first three questions forming the foundation of the fourth and final question. Answering these questions forms a chapter in their own right and the time of writing, two of the four chapters have been published with plans to publish the final two. As such the specific details of the literature reviews, the methods employed, the participants, tasks and measures taken are covered in corresponding chapters

4.7 Discussion and Conclusion

This chapter outlined the methodology employed to research, design, develop and evaluation model-based information navigation. The system will be constructed using open source web-based technologies such that they can be adapted to suit as well as maximising the interoperability between existing systems and file formats. The approach taken is broken down into four research questions, the first three address how the system should operate, covering the methods of document indexing, the techniques for navigating the model-based virtual environment and the methods by which information should be displayed. Each of these is then brought together to form the final test bed system that is used to evaluate the whole approach to model-based information navigation.

The next few sections now explain the actual detail of answering each of the research questions. **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing** answers Research Question 1: What are the most appropriate techniques for a model-based approach to document indexing? **Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments** answers Research Question 2: What are the most appropriate techniques for navigation information within a model-based virtual environment? **Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation** answers Research Question 3: What are the most appropriate techniques for displaying information within the model-based virtual environment? **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** and **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results** answers Research Question 4: How does model-based information navigation improve engineering information access and knowledge discovery?

Chapter 5

Appropriate Techniques for a Model-Based Approach to Document Indexing

5.1 Introduction

As touched on in **Chapter 3: Literature Review**, search engines are a component in the dissemination of information and knowledge and are implemented in a range of knowledge management and engineering information management systems. The Internet has seen technology mature and settle to the point where one rarely questions the effectiveness of the major search providers such as Google. This can not be said for Intranet or enterprise search however, the field has yet to see the big breakthroughs such as PageRank [59], that saw Google surpass competitors and cement its place in the market over the last 20 years or so.

The literature review concluded that there was a gap between the desired performance and that currently being delivered, even though research is delivering improvements through expanding search queries through context (ontologies, taxonomies and syn-sets). The emergence of model-based definition and building information modelling, where the physical and functional nature of the product is used as a means of accessing specific pieces of information and data, mirrors this concept of context - it being provided by the product structure itself. There is then sound logic for the combination of these two types of technology, search engines and visual representations of the product.

The first research question aims to address the first challenge in a framework for building such a system: *What are the most appropriate techniques for a model-based approach to document indexing?* That is, how do documents relate to the product structure? To determine this, this chapter beings with a literature review of the state-of-the-art of search technology and shows how the actual foundations of search have not changed for some time. Following this, and in answer to Research Question 1, the appropriate techniques of a model-based approach to document indexing are identified.

5.2 Background

This section will now present a background to the management of engineering information, an in depth review of the way search engines operate and the measures by which they are evaluated.

5.2.1 Shared Folder Structures

Prior to discussing the research fields of search engines and document classification, is has to be noted that the dissemination of information and knowledge through a knowledge/engineering information management system is not always feasible (due to lack of resources for example). In these cases, projects will frequently turn to more traditional shared folder on shared network drives for the dissemination of project documents. Figure 5.1 is an example from the University of Bath's Formula Student team from 2014/15. The folders are structured hierarchically such that each folder corresponds to branch in the product structure. Users find information by navigating to the folder relating to the area of the car that they expect to find the document.

Within the context of model-based information navigation, one can envisage a system built on the same form of structure with users opening three-dimensional components rather than folder icons. In this case, linking documents to the product structure would involve engineers manually making those connections, in the same way as the Bath Formula Student Engineers from 2014/15 did. At this point in this chapter, we can then say that a manual approach to indexing documents against the product structure is not only an option but it is one that has been seen to work within the confines of a Formula Student Racing team.

Figure 5.1: An example of the shared folder structure for the University of Bristol Formula Student Team

5.2.2 Search Engines

Whether Internet or Intranet the basic fundamentals of search are similar. The greatest challenge in finding information within a collection of documents stems back to library cataloguing [81] and while technology has digitised the process and quickened search, the processes have remained largely unchanged. This section explains these processes beginning with the indexing the corpus such that is it more accessible to the search engine. Following this the processes of expanding the search query to improve results relevance and ensuring the most appropriate documents appear first in a collection of results (ranking) are discussed.

5.2.3 Indexing

When performing a search using a search engine, the majority of 'search' is done well in advance of the actual typing a query into a text box and pressing return. The search engine does not trawl through every document looking for the occurrence of the query at the point of the user initiating the search. Every searchable document is processed in advance and it is the result of that pre-processing, a search index, that is used to find the relevant documents.

The underlying structure and methods of construction are well-established and as such, a textbook based on the teachings at Stanford University [82] is used as the basis of this subsection where a generic approach to search index building and in particular the Inverted Index is described. An index whose structure is optimised for the retrieval of documents.

Inverted Indexes are a particular type of index that are structured such that they are optimised for document retrieval (see Table 5.1). The index is formed of a dictionary and

Dictionary	Posting
term 1	doc ID 1, doc ID 2, ...
term 2	doc ID 4, doc ID 7, ...
term 3	doc ID 1, doc ID 3, ...
...	...

Table 5.1: The Inverted Index structure

posting. The dictionary being a list of terms that exist within the corpus and the posting being a lift of documents containing that term. A search involves matching the query to the corresponding term(s) in the dictionary and returning the list of document(s) listed in the posting. The search process itself involves moving down the dictionary and comparing the query against the dictionary term, one term at a time.

Recent examples from literature concerning the inverted index include the loading the index into flash memory [83], compressing the postings such that they are smaller hence quicker to move around [84] and partitioning the index over a number of servers [85] again to minimise the load of any one server in an attempt to improve performance. It is interesting to note that in each of these examples (and in the many more not discussed here) the fundamental Inverted Index structure is left unchanged and the experiments are focused on increasing the speed of interacting with the index. It is then possible to say that in constructing a means of information/document retrieval, the inverted index is the optimal choice.

The process of creating the index is relatively simple although there are a number of steps and options involved that can merge and increase complexity. The steps include, for example, how one manages spell-checking/correction, acronyms, synonyms, plurals, and lemmatization to merge word forms (engine, engineering and engineer). Regardless of each of these, the first step is to iterate through the corpus, one document at a time, extract each term (tokenization) before adding the instance of the term-document pair to the index.

Tokenization is the first step in traditional document indexing and involves splitting up text into individual words or phrases. In its simplest form this can be done by splitting the text using white spaces but consideration should be made for words ending in special characters e.g. -,.;)*. It is these individual terms that form the dictionary component of the index.

Stop words are words within the corpus that both occur too frequently and whose meaning do not add anything to differentiate between documents. Words like *and*, *it*, *some*, *the*, etc. can be ignored or removed from the dictionary.

Stemming encapsulates the meaning of a word while reducing it to the shortest set of characters that cover all possible word endings. For example the Porter Stemmer [86] shortens engine and engines into engin and the WordNet Lemmatizer [87] shortens engine and engines to engine and leaves engineer and engineering as engineer and engineering.

Spell checking handles those situations when the user enters a search query and the word does not exist in the dictionary or a document contains words that are misspelled. There are a number of techniques for correcting user spelling mistakes but the simplest will offer the user a choice of the next closest word in the index dictionary. Deciding which word is closest is a matter of navigating the dictionary but when multiple words of the same 'closeness' are returned, it is common for internet search engines to return the most frequently requested. In theory, suggesting the most frequent gives the best chances of determining the correct word.

Text normalization is the process used to overcome variations in how words can be written, *U.K*, *UK* and *United Kingdom*, *auto-pilot* and *autopilot*. The traditional approach for this is to use implicit word classes where all variants of the word are mapped to a common word (auto-pilot and autopilot being mapped to autopilot). Using this approach requires some care, as an example, should *C.A.T.* return matches for *cat*.

5.2.4 Expanding Search Queries

Language is complex and the elements that make it a powerful means of communication, its flexibility, results in it being difficult to computationally analyse for the purpose of indexing. There are then elements of the process of building an Inverted Index that do not necessarily mirror the means by which users search for information. Techniques for mitigating these elements tend to focus of expanding the search query such that the query's context is utilised. This subsection will now outline some of these techniques.

Syn-sets are the response to the synonym problem, when the user searches for either *plane*, *aeroplane*, or *aeroplane*, a syn-set strategy will have combined all these terms within the index such that the results are returned for all three terms rather than just the one used within the query. Syn-sets are of particular importance when working with enterprise search where often different department/functions/systems will have different names. Syn-sets allow for search across department/functions/systems using the terminology familiar to the user.

Taxonomies and Ontologies attempt to expand search using additional terms that are related to the query. The relationship is case dependant, an aerospace example would be expanding the query *wing* to also include parts of the *wing* - *aileron* and *flaps* for example. Taxonomies group words in a hierarchy of concepts (a golden retriever is a dog which belongs to the canine family of mammals). Ontologies capture relationships between concepts (a golden retriever eats dog food and walks in a park). Capturing the relationships between concepts allows for a more intelligent search and so the returning of more appropriate results.

Personalised search aims to achieve the same benefits of syn-sets and taxonomies/ontologies by leveraging the context of the user themselves. The technique relies on a profile of the user that is used to stir the search away from areas of dis-interest and towards those areas of interest. A car enthusiast searching with the query *jaguar* would more likely be shown results relating to the car company and not the animal, whereas a zoologist would more likely be shown results related to the animal and not the car. Profiles can be generated algorithmically and have been shown to improve research performance [88].

5.2.5 Results Ranking

Ranking the results of a search can be as equally as important as the indexing process itself. If the most relevant results are not shown towards the top of a list of results, then users face trawling trough the entire list to find the desired information/knowledge. It was Google's PageRank ranking algorithm that lead to it becoming the leading search engine that it is today. Ranking is simply the order in which the list of search results are displayed to the user. What is not simple however is the governance of that order. This subsection now explores ranking, covering some of the most popular and relevant/irrelevant techniques for enterprise search.

PageRank [59] was Google's ground breaking ranking approach that pushed Google to become the dominant Internet search engine. The approach listed results by the number of other web pages that linked to it. The higher the number of links, the higher a page appears in the ranking. The approach is successful in that it captures Internet content creators (users/web designers who post content on web sites) views on web pages - content creators create links to pages that they deem worthy of linking to. The more links than, the more content creators have evaluated and concluded the good content.

While this works on the Internet, it does not work in enterprise search. There are no equivalents to hyper-links in most business file types (word documents, spreadsheets, CAD files, source code, etc.) and such the measure of number of links does not exist. This also assumes that the number of links to a business document would be an appropriate ranking technique if the links did exist.

Popularity simply counts the number of times users click-through to a particular document [89]. The more users click on a particular document the higher up the ranking that document appears. This technique does go some way to capture the user-determined appropriateness of a given document to a particular search query however it is not without faults. For example, in a given search, if a user clicks on five of the top results before ending the search, should all five of those clicks count towards future ranking or should the system assume that the final click satisfied the information need and so only that click should count.

There are also examples from literature that show how the popularity of a document may not be appropriate. Certain business documents are periodically published, finance reports for example, ranking by popularity will always result in older documents being ranked higher, given the accumulative click counts over time. There is literature that discusses cases where users complain about this particular behaviour [9].

Date/Time, Author, Customer Review, Version, Status, Size, etc are examples of other measures to rank against that are not necessarily widespread in general Internet search but do appear in smaller domain specific applications, eBay [1], Amazon [2] and file system search (see Figure 5.2 - files are ordered by clicking on the Name/Size/-Type/Location headers) to name three examples. One can argue that certain measures are more appropriate for certain business documents/file types/circumstances. For example, if one were searching for a particular CAD file, is it more appropriate to see the latest version or the latest 'stable' version?

Faceted Search is a technique that classifies results into particular classes (or facets) and allows users to switch between and classes to filter results. Examples of these are

[1] `https://www.ebay.co.uk/`
[2] `https://www.amazon.co.uk`

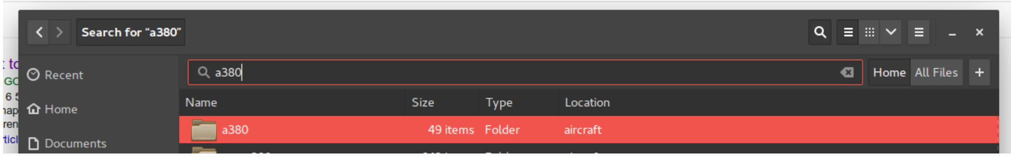


Figure 5.2: Search within the Nautilus file manager from the Linux Gnome desktop environment

	term 1	term 2	term 3	...
document 1	1	1	1	...
document 2	1	0	0	...
...	...	...	...	...
query	0	1	1	...

Table 5.2: Corpus documents and a search query represented in a Vector Space Model

found on domain specific search websites on the Internet. Autotrader.co.uk[3] is a second hand car sales specialist and is one example. Users can search by both keyword and/or selecting values from pre-defined classes from lists (make, model, age, mileage bracket, distance away, for examples). Examples of faceted search from engineering literature include the organisation of design documents [90]. Faceted search has the advantage of dramatically narrowing the search space although this comes at the cost of a greater level of user input.

Cosine Similarity is a ranking approach that linguistically generates a measure of comparison between documents against the search query. Results are then ranked by linguistic similarity. The first step involves re-representing the corpus within a **vector space model** [91]. The vector space model represents a document/search query in multidimensional space, with each dimension corresponds to a term within the entire corpus (see Table 5.2). Documents/the query are then vectors in that space where the components have a value of 1 or 0 based on whether it contains a particular term. In its simplest form the similarity measure between two documents is then the cosine distance between the two vectors (see Figure 5.3). This type of analysis is also known as a Bag-of-Words analysis. A search engine can turn a user's query into a vector within the vector space model and generate a similarity score for every document. This score can then be used to rank the results when they are returned.

Term Frequency - Inverted Document Frequency (TF-IDF) is a corpus linguistics approach for measuring the importance of terms within a corpus (D) and is

[3]`https://www.autotrader.co.uk/`

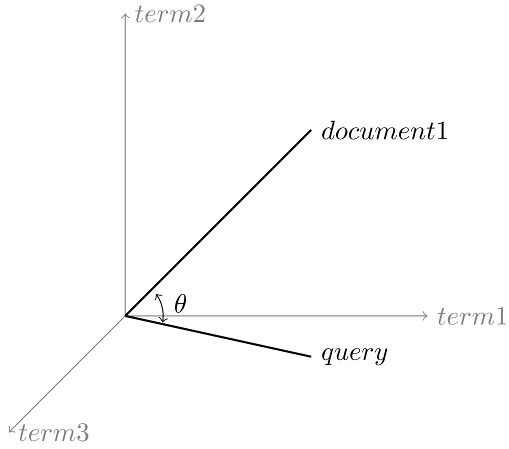


Figure 5.3: Cosine Similarity (θ) within a Vector Space Model

widely used throughout information retrieval and |colorreddocument classification. It is a combination of two measures, the Term Frequency (tf) and the Inverse Document Frequency (idf). The term frequency is a count of the occurrence of term t in document d (Equation 5.1). The inverse document frequency is the natural log of the total number of documents (N) divided by the number of documents containing the term t (Equation 5.2). The tf-idf is the dot product of the two measures (Equation 5.3).

$$tf(t,d) = f_{t,d} \tag{5.1}$$

$$idf(t,D) = \log\left(\frac{N}{\mid (d \in D : t \in d) \mid}\right) \tag{5.2}$$

$$tf\text{-}idf(t,d,D) = tf(t,d) \cdot idf(t,D) \tag{5.3}$$

Ranking results by the tf-idf is a matter of extracting those documents containing the term(s) and comparing the term weights. Those most similar to the query are ranked higher. Again the cosine distance can be used to determine a similarity measure between the document and query however in this case the term weights are using in place of the binary 1 (contains term) or 0 (does not contain term).

Evaluation: Precision, Recall, and F1-Score

A traditional approach to evaluate search engines are the measures of *precision*, *recall* and $f1$-*score* [80]. Precision is the number of relevant results returned divided by the total number of retrieved results (Equation 5.4). Maximum precision would be a system that returned every correct result in the corpus and none of the incorrect results. In reality, maximum precision is rarely achieved. Recall is then the number of relevant results returned divided by the number of relevant results that should have been returned (ground truth) (Equation 5.5). To obtain a better understanding of a system's effectiveness it is important that these two measures be used together and the f1-score or f-measure combines the two to give a single measure (Equation 5.6).

$$precision = \frac{\mid \{relevantdocuments\} \cap \{retrieveddocuments\} \mid}{\mid \{retrieveddocuments\} \mid} \tag{5.4}$$

$$recall = \frac{\mid \{relevantdocuments\} \cap \{retrieveddocuments\} \mid}{\mid \{relevantdocuments\} \mid} \tag{5.5}$$

$$f1\text{-}score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5.6}$$

5.2.6 Summary

This section has provided a general overview of search engines, their construction, their operation and their evaluation. There are entire books and theses written on the field and so while there is a vast amount of information that is not covered here, this section has been written to provide a theoretical background and underpinning for the answering of the first research question. The reminder of this chapter explores and answers this question.

5.2.7 Machine Learning and Document Classification

Tan et al. [92] defines classification as "*...the task of learning a target function that maps each attribute set x to one of the predefined class labels y*". An example of classification in practice is applying for a loan from a bank. The decision of whether the loan is offered or not is made based on the likelihood of the applicant repaying the loan. This likelihood is determined based on a number of attributes like age, sex, location, income, employment status, marital status, repayment history, number of other loans and so on. When an application is made, the bank enters the applicant's details into a model and the output of the model will place the applicant into a class that predicts whether the loan will be repaid.

Generating the model requires training data, a pre-classified data set containing all the attributes (age, sex, location, etc.) along with whether the application repaid their loan (the class). The model in generated by finding patterns in the training data that predict the output (able to repay or not). Predicting whether a new application is likely of repaying or not is then a matter of matching their attributes with those similar existing cases in the model.

Document classification is then the process of automatically classifying documents against a pre-determined list of classes. This is important within the context of this thesis given the model-based environment has a finite number of components with which the user can interact with to access documents, unlike a text-based search engine when the user can enter any string of text. If the number of components is finite then the indexing problem can be considered as a classification problem, which opens the document indexing to the world of document classification.

Case-based reasoning, is an example of classification where new documents are classified against a set of classes based on the documents that already exist within each class. The main difference between this approach and general classification is the real data is used for every new case rather than a model generated from existing data. For example, imagine one has a number of folders relating to the engine, chassis, suspension and steering of a car. Each folder contains a number of documents relating to that particular part of a car. A new document is created and needs to be automatically saved in the correct folder. Case-based reasoning would compare the new document against all the other documents in the all the folders and select the folder with the most similar documents. That comparison can be achieved through the techniques such cosine distance or TF-IDF discussed in the search indexing section of this chapter.

5.3 Appropriate Techniques for Model-Based Document Indexing

Until this point this chapter has focused on a generalised approach to building search engines in an attempt to give a theoretical foundation for answering Research Question 1: *What are the most appropriate techniques for a model-based approach to document indexing?* The foundation of search engines is based on documents being searchable by any term within a corpus. However, the model-based approach to indexing however, leans towards faceted search in that if users are clicking on specific system/sub-system/component within the product structure, then each system/sub-system/component can be considered a class against which, documents can be classified. Classification is a process used to algorithmically label a data set against a pre-set list of classes [80]. The question then is, can document classification be used to classify documents against the product structure and improve engineering document search compared to a more traditional search engine? Is this an affordance of model-based information navigation?

Attempts to make improvements in the field of information retrieval have examined techniques such as syn-sets and ontologies/taxonomies. These attempt to improve search by using relationships and concepts between terms. Ontologies/taxonomies capture the relationships between terms and concepts and use these to expand and filter searches. A search for *powertrain* could be expanded to return the results for all the *powertrain* components. Synsets expand the search query by including the terms with the same meaning, for example *powertrain* and *engine*. These techniques use structure to expand the query such that the results include a wider range of related terms. For example, a document that contains multiple terms related to and including *powertrain* or those that have the same meaning, are more relevant to the user than those documents that do not include those terms.

The hierarchical tree structure of the product structure is part the ontology/taxonomy that these techniques aim to leverage and as such, using the structure should inherently improve document retrieval. While the classification could be achieved manually, this would be a laborious process for large collections of legacy document. Hence, the need for an automated process. The remainder of this chapter describes an investigation into developing such an approach and its evaluation compared to a traditional TF-IDF search engine. The purpose of this study is to determine whether or not classification against the product structure can be used to improve search over more traditional methods, TF-IDF being an example of a more traditional Search Engine.

5.3.1 Indexing Method

Figures 5.4 and 5.5 show high level process diagrams for creating a traditional TF-IDF search engine and the classification approach respectively. The traditional approach analyses documents using the Equations 5.1, 5.2 and 5.3 to generate the TF-IDF term weights and stores them in an inverted index as shown in Table 5.1. The TF-IDF search engine generates a matrix that represents each document in the corpus as a list of terms and a corresponding TF-IDF weight that reflects the importance of that term to that document within the context of the corpus. Searching the TF-IDF structure with a search query involves retrieving each document containing a non-zero TF-IDF weight for term(s) contained within the query. The retrieved documents are then ranked based on the weight/summed weights with the highest appearing at the top of the results list.

The classification approach has several additional steps to the traditional approach however these breakdowns into relatively simple stages. Figures 5.4 and 5.5 describes

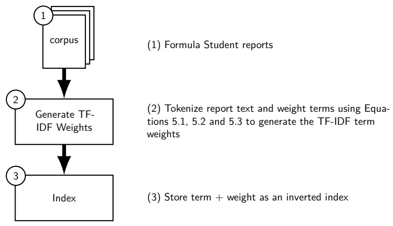


Figure 5.4: The process model for a traditional TF-IDF search engine

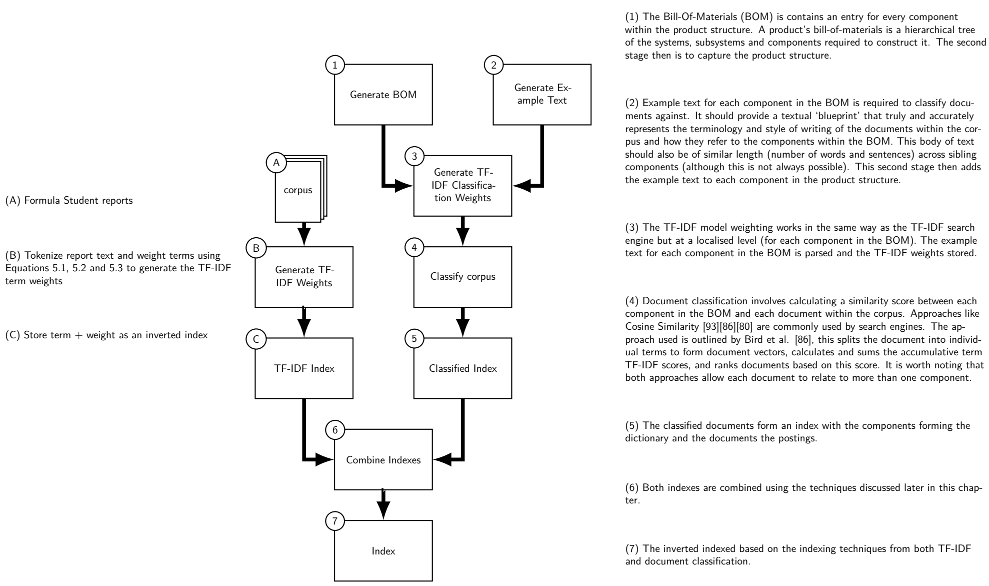


Figure 5.5: The process diagram for the combined classification and TF-IDF approach used for document indexing

each of the stages in the classification approach used.

5.3.2 Data Set and Implementation

The Institution of Mechanical Engineers Formula Student is a competition that requires university students to design, build and race a single seat racing car. Teams of around 30 students design the car in their third year and construct it in their fourth year. For the purposes of this study, the 2013-14, 2014-15 and 2015-16 reports from the University of Bath were used. Reports were in PDF format, around ten pages in length and consists of raw unstructured text, tables and images. In total this corpus is comprised of 281 reports. Table 5.3 shows a more detailed description of the textual content of the corpus.

Figure 5.4 shows the process diagram for creating a traditional TF-IDF search engine, and Figure 5.5 shows the combined approach. Figure 5.5 stages 1 and 2 show the construction of the TF-IDF search engine. This was achieved by extracting the text

Statistic	Number
Documents	281
Average words per document	4671
Average unique words per document	724

Table 5.3: A number of statistics describing the data set used

Level 1	Level 2	Level 3	Level 4
fs car	engine and drivetrain	exhaust system	secondary pipe
fs car	engine and drivetrain	exhaust system	muffler
fs car	engine and drivetrain	oil system	oil tank de-aerator
fs car	engine and drivetrain	oil system	overflow bottle holder
fs car	engine and drivetrain	oil system	overflow bottle
fs car	engine and drivetrain	oil system	oil coolant heat exchanger
fs car	engine and drivetrain	fuel system	fuel tank
fs car	engine and drivetrain	fuel system	filler neck hose
fs car	engine and drivetrain	fuel system	filler neck hose mouth
fs car	engine and drivetrain	fuel system	filler neck cap
fs car	engine and drivetrain	fuel system	drain cap
fs car	engine and drivetrain	cooling system	water radiator
fs car	engine and drivetrain	drivetrain assembly	chain guard
fs car	engine and drivetrain	drivetrain assembly	differential assy
fs car	engine and drivetrain	drivetrain assembly	differential carrier

Table 5.4: A section of the Bill-Of-Materials used for classification

from all 281 reports and generating the TF-IDF weights for each term in the corpus. These were then stored in an index that was interrogated to produce the combined index and the TF-IDF results for the comparison.

In addition to the reports, students submit a financial statement which includes a Bill-Of-Materials (BOM). Each year students generate a new design and with it a new BOM. For the purposes of this study a generic BOM was created using the most common components and component names from across the three years. Table 5.4 shows an extract from this generic BOM. This relates to stage 1 in the process diagram shown in Figure 5.4.

The next stage (2) involves obtaining component level example text. This was achieved by asking a domain expert (J) to generate bodies of text for each component in the BOM. The approach to use a domain expert was taken as there was not a body of suitable documents available to act as example texts and the expert was deemed able to provide suitable alternative. J was selected due to their 30 years' engineering experience with 10 years specifically in an academic leadership role within Formula Student.

The study was explained in full to J, including the document classification approach taken. J was given the brief of generating text descriptions for each of the components within the BOM used. The descriptions were requested to be of similar length and level

Part	Example Text
brake kill switch	This is an over travel switch from part of the brake pedal assembly. If the brake pedal over travels, the switch will be activated and will stop the engine from running by 'killing' the ignition and cutting the power to any electrical fuel pumps.
brake light housing	The 'brake light housing' is the structure that retains the lens, bulb and wiring of the brake light.
brake light	The brake light is a warning device on the rear of the vehicle that eliminates when the driver presses the brake pedal.
brake pad	Brake pads are a component in the braking system of the vehicle. Their general construction is of a steel backing plate with friction material bonded to the surface. They are generally fitted as a pair in a calliper sitting either side of the brake disc, either the friction material touching the disc surface. When the driver presses the brake pedal hydraulically force is applied, the calliper clamps or squeezes the two pads together into the spinning disc to slow/stop the vehicle.
brake pedal assy	The brake pedal assembly is a collective term to cover all components that make up the brake pedal unit.
brake pedal bracket	The brake pedal bracket is a fixture that retains the brake pedal assembly to a required position in relation to the driver and the chassis.
brake pedal rod end	These are part of the brake system. These are eyelets that connect the brake pedal to the linkage arm. Commonly in a clevis or spherical bearing design.
brake pedal shim	Brake pedal shims are thin profiled pieces of material designed to change the clearance or distance of parts of the braking system that are assembled.
brake pedal	The brake pedal is a foot operated lever used by the driver of a vehicle to operate the brakes. It consists of three parts: arm, pivot attachments and pad.
brake system assembly	Is a collective term for all elements that make up the braking system.

Table 5.5: An extract from the BOM and associated example text

of detail across all components, although some components with more complex than others and as such required more text. The level of detail (text containing all relevant information and hence all relevant terms) was prioritised over having text of similar length to allow for these varying levels of complexity. J was also asked that the text be representative of the manner in which components are discussed in Formula Student reports. An extract from the result is shown in Table 5.5. On average, each example text contained 31 words split over two sentences.

Stage 3 constructed the model by parsing each component's example text and performing a TF-IDF comparison with its siblings' components example text. TF-IDF scores were then stored for each component. An extract from the model for the component 'Crank Sensor' is shown in Table 5.6.

Stage 4 includes the classification of the FS reports. This was done by extracting raw text from the PDF reports and tokenizing the text to obtain a list of terms in each report. For each component in the BOM model the TF-IDF weight for any shared term was summed to give each document a similarity score.

Part	Term	TF-IDF Weight
1	crank	0.0084170581
	speed	0.0084170581
	crankshaft	0.0084170581
	combustion	0.0068013304
	engine	0.0068013304
	internal	0.0068013304
	rotational	0.0058561903
	monitor	0.0058561903
	sensor	0.0051856027

Table 5.6: An extract from the TF-IDF weight for terms for a BOM component example text

By its nature, the model generates a similarity score for all documents within the corpus. This is counter-productive in IR systems given the system could return every document in the corpus when a search is performed. The only difference between searches is therefore in the order that results are returned. Hence, there is a need to restrict the number results returned by the classification system to those most relevant. The localised (component and system) level weighting of terms within the BOM structure means weight cannot be compared across the entire product structure and so a simple generic threshold score could not be universally implemented. The technique used is explored and discussed in the results section. In addition to the cut off, the Results section also discusses techniques for combining the two sets of results. Essentially, the Results section of this chapter discuss techniques for delivering an effective Stage 7 (Figure 5.5), the merging of the two indexes. Stages 1 and 2 show the construction of the TF-IDF search engine using the exact same method as shown in Figure 5.5.

A search involves traversing both the TF-IDF search index and BOM classification search index. The results from both are then combined and ranked based on the summation of the two scores and this is discussed in the following sections. For the purposes of the study and to explore the potential benefits, seven component names were selected at random from the BOM.

5.3.3 Results

This section focuses on two main areas. The first is a strategy to limit the number of results returned by the classification system. The second examines two approaches to combine the results from the two techniques. In order to evaluate the search results from the classification approach and the traditional TF-IDF, searches were performed on seven terms/components selected at random from the BOM (shown in Table 5.7). For

Query	Ground Truth	Retrieved Relevant Documents	
		Classification	TF-IDF
Brake System	38	22	19
Cooling system	21	17	13
Exhaust System	40	30	19
Front Wing Assembly	26	21	7
Paint - Body	2	0	0
Steering Column	8	6	5
Track Rod	3	0	1

Table 5.7: The number of retrieved relevant documents returned by each approach

Query	Ground Truth	Total Retrieved Documents	
		Classification	TF-IDF
Brake System	38	229	86
Cooling system	21	232	81
Exhaust System	40	240	66
Front Wing Assembly	26	238	45
Paint - Body	2	0	0
Steering Column	8	238	22
Track Rod	3	0	48

Table 5.8: The total number of retrieved documents returned by each approach

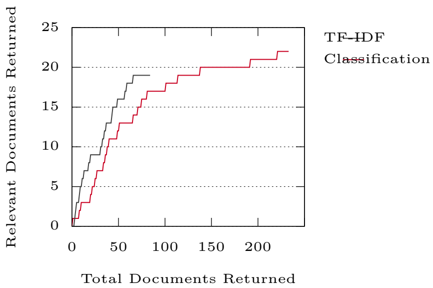
comparison, the ground truth for the seven terms was generated by the same domain expert who generated the example texts. From the seven terms used, five returned results for both approaches, see Table 5.7. Neither approach returned 100% of the ground truth documents. The classification approach returned a higher number of relevant results however, as expected, Table 5.8 shows how the approach also returns a far higher number of non-relevant results.

Cut-Off

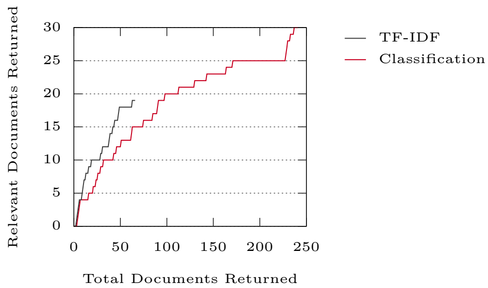
Several techniques were tested to find a representative cut-off for the classification results, top n-results or top x-percent, for example. Further study in this area is needed and not covered by this thesis given the aim here is to show affordance are possible using classification rather than to optimise said affordance. However, the closest representative measure found was using the number of non-zero results returned by the TF-IDF approach, i.e., those containing a TF-IDF weight greater than zero.

Figure 5.6a, Figure 5.6b, Figure 5.6c, Figure 5.6d, and Figure 5.6e show the number of relevant documents returned versus the total number of documents returned for the 'Brake System', 'Exhaust System', 'Cooling System', 'Front Wing Assembly', and 'Steering Column' respectfully. As shown in Table 5.8, no results were returned for the terms 'Paint - Body' and 'Track Rod' and so those figures have been omitted. All five Figures (5.6a, 5.6b, 5.6c, 5.6d, and 5.6e) present a similar trend, the classification approach classifies most documents with a non-zero score and as such most results are returned and in comparison, the TF-IDF approach returns far fewer non-zero results.

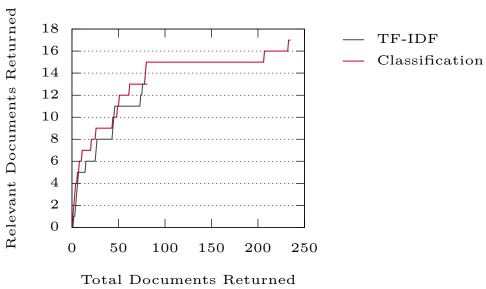
Within each of these figures, a steeper slope represents a higher precision as more relevant results are returned earlier in the total number of results returned. Four of the figures (5.6a, 5.6b, 5.6d, and 5.6e) show how the TF-IDF performs with better precision with the Figure 5.6c showing more mixed results with the classification approach preforming slightly better. From this, it was decided that the number of results returned by the TF-IDF approach can be used as an appropriate technique for limiting the total number of results returned by the classification approach such that the results from both sets can be combined.



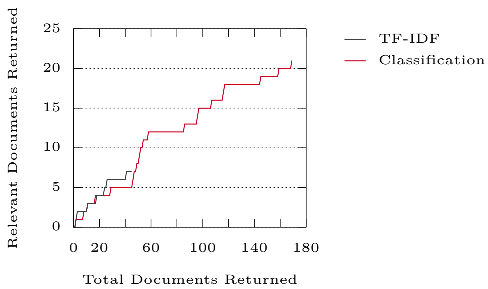
(a) Brake System



(b) Exhaust System



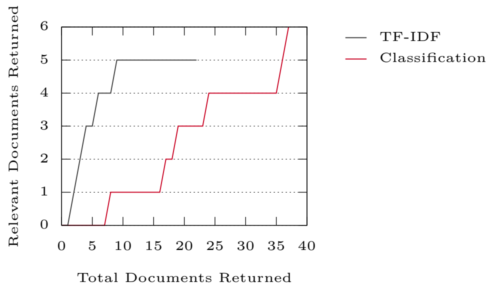
(c) Cooling System



(d) Front Wing Assembly

Combining Results

Results were combined using two approaches: merge and intersect. Merging results involves combining the results in order one at a time before removing the duplicates. When removing duplicates, the highest ranks results was kept. The intersect filters only those results that appear in both approaches and again the highest rank for each result was kept.



(e) Steering Column

Figure 5.6: Relevant versus Total Returned Results

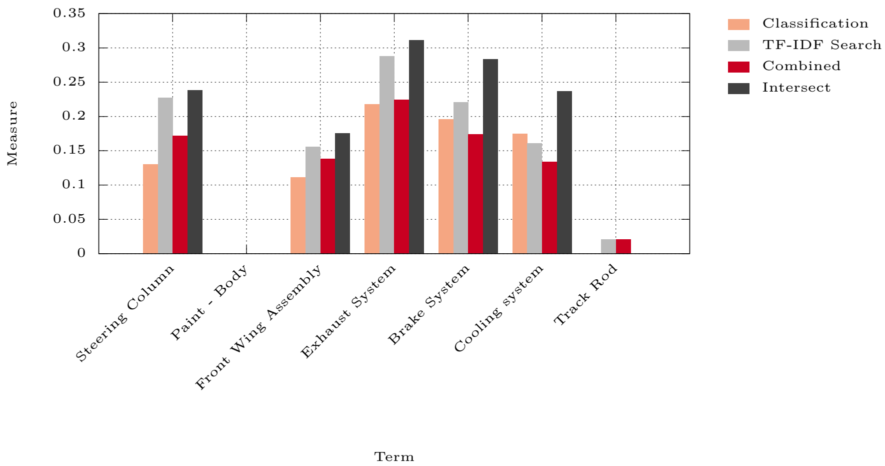


Figure 5.7: The Precision values for the four approached

Figures 5.7, 5.8, and 5.9 show the precision, recall and f1-score for the classification, TF-IDF, merged, and intersected results. These figures show a mixed and relatively poor performance of across all measures and terms searched, with the recall performing better than the precision. However, Figure 5.7 shows how the intersect of the classification and TF-IDF results improved the precision across all five terms with results for both classification and TF-IDF. Figure 5.8 show the recall is improved by merging results. Figure 5.10 and 5.11 show the percentage improvement for each term for the merged and intersected results respectfully.

The first noticeable difference is how the merged results boost the recall while reducing the precision and f1-score. The opposite is true of the intersect where the precision and f1-score are increased while a drop in the recall occurs for two terms and no change is seen in the other five terms. Both Figures 5.10 and 5.11, do however show a result that goes against these trends. The large percentage improvement for the 'Front Wing Assembly' causes a very slight increase in the f1-score. Looking at the figure for the

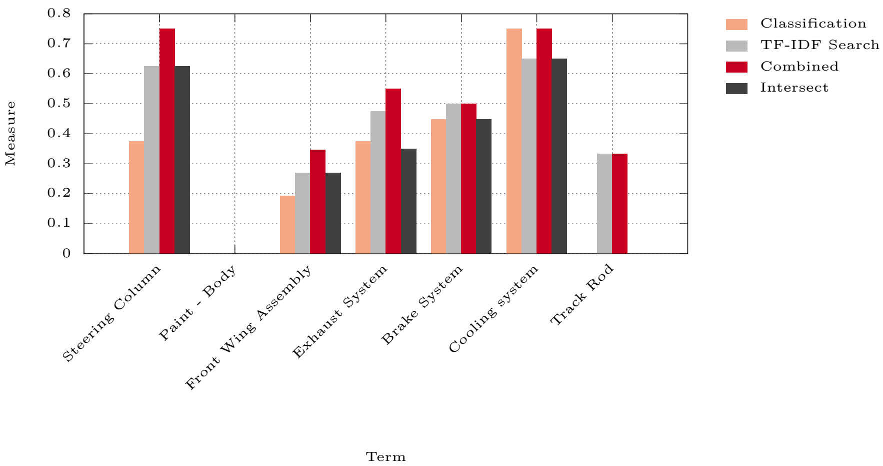


Figure 5.8: The Recall values for the four approached

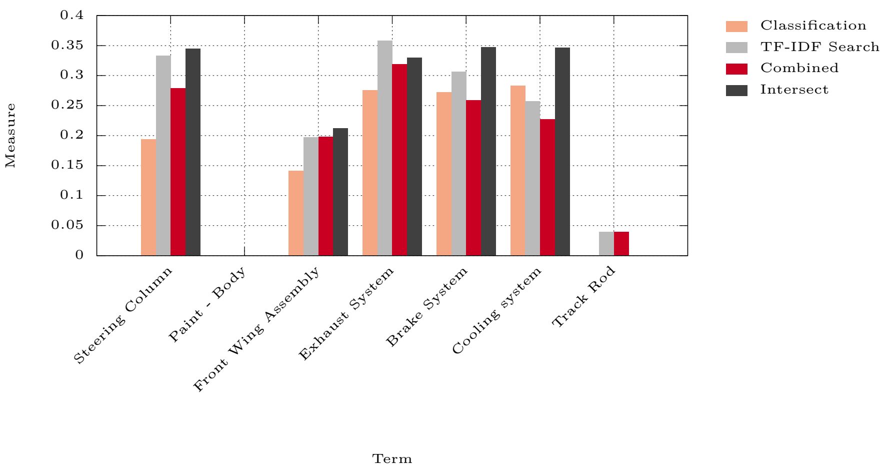


Figure 5.9: The f1-score values for the four approached

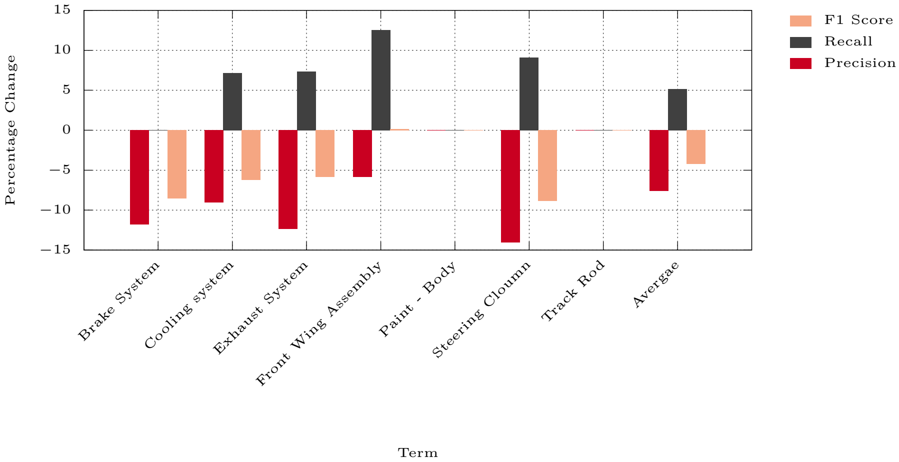


Figure 5.10: Brake System: Relevant versus Total Returned Results for the Merged Set of Results

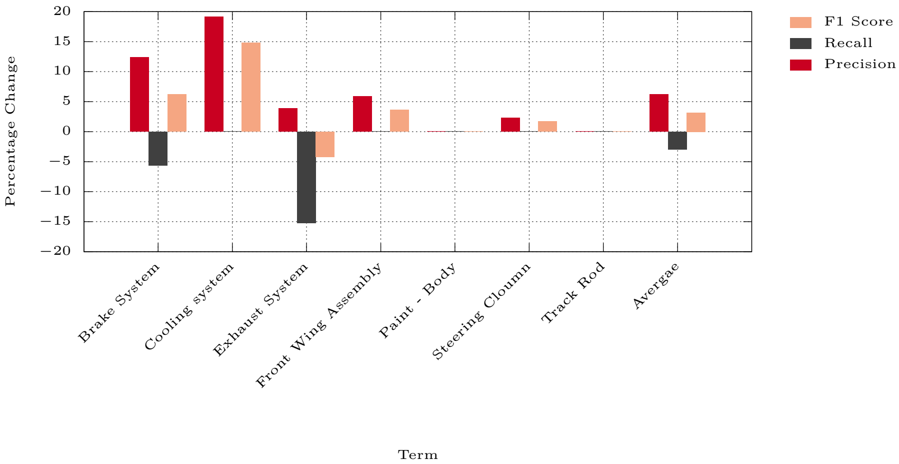
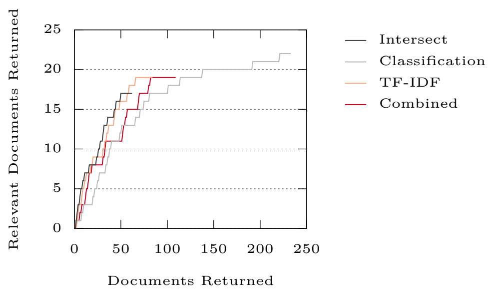


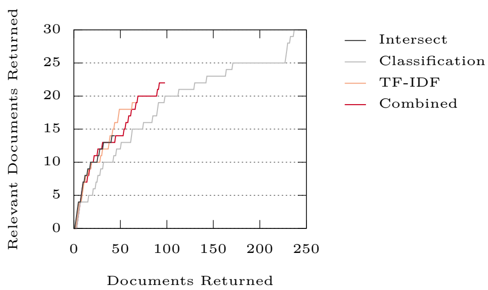
Figure 5.11: Exhaust System: Relevant versus Total Returned Results for the Intersected set of Results

intersect, a large percentage decrease in the ‘Exhaust System’ also causes a decrease in the f1-Score.

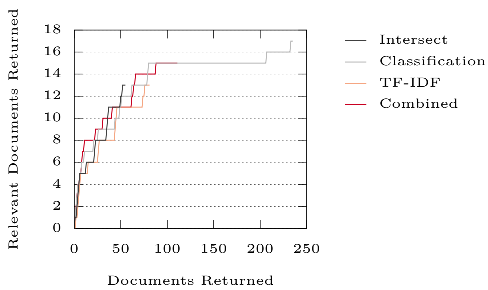
Figures 5.12a to 5.12e show the same data as Figures 5.6a to 5.6e with the addition of the values for the merged and intersected results. The results for the intersect approach again show how precision in improved over the other approaches, with more relevant results being returned sooner. The results for the recall show how the merge approach falls in between the classification and TF-IDF approached, with a slower return of relevant results, but more relevant results being returned over all when compared to the intersect and in four out of five when compared to the TF-IDF results.



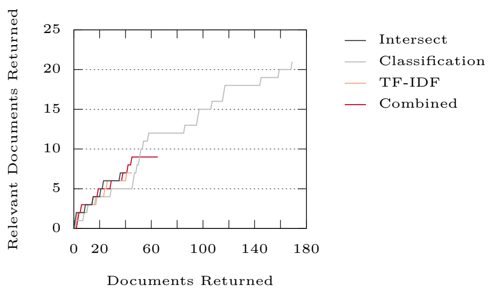
(a) Brake System



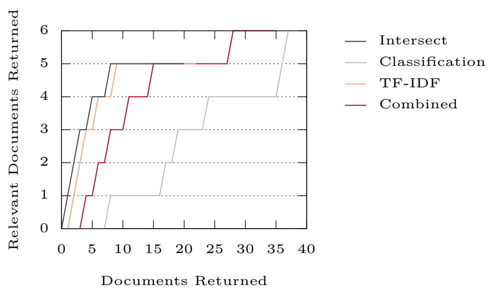
(b) Exhaust System



(c) Cooling System



(d) Front Wing Assembly



(e) Steering Column

Figure 5.12: Relevant versus Total Returned Results

5.3.4 Summary

The approach outlined in this section aimed to improve a TF-IDF search system by combining the TF-IDF with a classification system that classified documents against the product structure. To investigate this concept a search engine was constructed that generated results for the two approaches and methods for combining these results were examined. The results show that the number of results returned by a TF-IDF search generated a representative cut-off point for the number of results returned by the classification approach used. In combining the results from the two approaches, promise was seen depending on whether the goal is to expand on the number of relevant results returned or to increase the precision of those results.

The goal of any search engine is 100% precision but challenges such as the ambiguity of language and the uncertainty in the user's information needs mean that there are no perfect IR systems and all searches are carried out with the expectation that the final stages of the search will be performed by our own evaluation and browsing of a corpus subset. The work presented here shows that the technique provides some tailoring of this subset. An example its usefulness was presented by Jones et al. [74] where IR is performed via a three-dimensional visual representation of the artefact. Representing many results within a three-dimensional artefact space will quickly clutter the visualisation and become unusable. In this case the intersect of the two result sets will benefit the visualisation, and the possible reduction in the recall may be acceptable given an increase in precision and greatly reduced number of documents returned.

There are however several areas that warrant further research. The size of the corpus used here (281 documents) did not allow for the more traditional division of the corpus into training and testing sets and so example text was generated and used. One can question whether the example text is representative of the corpus itself and it would be beneficial to repeat this study on a larger corpus, for example those used within large organisations such as Aerospace where mature products generate larger corpora, standardised lexicons, reporting/documentations and procedures. It is worth noting, however, that the corpus used is a 'real-world' example and while results may improve with a larger data set (thousands or tens of thousands of documents), research will at some point need to deliver solutions for these 'real-world' challenges of small corpora.

There are several alternative document classification and machine learning approaches to construct the classification model, for example artificial neural networks and deep learning. The aim of the work presented here was to determine if the product structure can be used to improve IR and not to determine the best approach for doing so. Now that it has been shown that the technique can have a positive impact on the results returned the foundations are in place for this future work.

The final aspect to discuss is whether there are better strategies for merging the two results sets (compared to the intersect and merge approaches presented). This study showed the intersect improved precision while the merging improved the recall and f1-score. No attempt was made to integrate the two approaches and generate a result set that optimised all three measures. This would also benefit from further study.

5.4 Discussion and Conclusion

This chapter aimed to answer the Research Question 1: *What are the most appropriate techniques for a model-based approach to document indexing?* Through a literature review and investigation into the use of document classification as a means of improving document indexing, the findings of this chapter are:

- That manual indexing is an option for linked documents to the product structure (although this can be laborious for large data sets)
- Document classification techniques can be used to improve document indexing against the product structure when compared to a more traditional TF-IDF document indexing approach
- Document classification techniques can improve the precision of a traditional TF-IDF search engine which is beneficial for displaying search results graphical representations as it reduces visual clutter

There is however a lot more research that this topic would benefit from: improved classification/machine learning techniques and a formal mathematical description of the work for example. However, one could argue that these would amount to a thesis in its own right and this thesis has set out to explore the wider aspects of delivering a model-based approach to information navigation and the associated affordance and best practices. This indexing proportion of the thesis is then concluded here and the next two chapters move on to identify the navigation and information display techniques, or answers to research questions two and three.

Chapter 6

Navigation in Model-Based Three-Dimensional Engineering Virtual Environments

6.1 Introduction

Finding information in three-dimensional virtual environments is a problem largely rooted in the field of human-computer interaction. Human computer interaction is the research area concerned with the barrier between the user and the computer system and covers areas from psychology of user behaviour to hardware design (mouse, keyboard, virtual reality headset) to what the user sees, hears and/or feels and how the system responds to the user's actions [94]. The early days of user interface research generated the Windows, Icons, Menus and Pointers that are still the basis by which we interact with desktop and laptop computers [95][96]. More modern outputs include the realisation and commercialisation of voice control [97] in Google and Apple Amazon's personal assistants [98][99].

Virtual environments are defined by [100] as "*A synthetic, spatial (usually three-dimensional) world seen from a first-person point of view. The view in a virtual environment is under the real-time control of the user.*". Engineering virtual environments are then a subset of virtual environments developed with an engineering application in mind, Computer Aided Design (CAD) software being the obvious example. These environments allow engineers to interact with engineering artefacts and those interactions can include drawing (CAD), simulation (Player Stage[101]), control system design (Simulink [102]), testing (finite element analysis[1] and computational fluid dynamics[2]), training (flight simulation[3]) and the information navigation system proposed in [74].

Interactions in wider three-dimensional virtual environments is a well established field that divides this interaction into three components: navigation, selection and manipulation, and system control[94][100]. Any good virtual environment will have mastered each of these to achieve an intuitive and efficient user interface and the system proposed by this thesis is no different. While traditional CAD packages provide interfaces based on sound theory and decades of user feedback, the task of drawing and designing an engineering artefact is different to that of finding information within an engineering artefact space and so methods can not be simply transferred between the use-cases.

Furthermore, at the time of writing, there exist no guidelines for designing systems to aid the finding of information in three-dimensional virtual environments. Design theory can suggest techniques for highlighting information objects that make the markers more visible to the user however, if the marker is not visible in the user's current view strategies are needed to manage how the users navigate to a view where the marker is visible. The field of human-computer interaction can be used to identify best practice and exploring the state-of-the-art and Information Foraging Theory [103] produces insights into how users naturally find information which can be used to avoid implementing a system that might inhibit users from "doing what comes naturally" when searching.

Users have been finding information in computer games for as long as computer games have existed and while Link walking around cutting down bushes in the Legend of Zelda may not seem applicable, the games use of "world-view" maps and waypoints could be. It is then worthwhile to explore fields that search within three-dimensional environments and this chapter now explores a range of applicable virtual environments and theories to develop a navigation strategy that is appropriate to support engineers who are interacting with three-dimensional engineering virtual environments. The aim of this chapter then, is to answer Research Question 2: *What are the most appropriate techniques for navigation information within a model-based virtual environment?*

[1] `https://www.autodesk.co.uk/solutions/finite-element-analysis`
[2] `https://www.autodesk.co.uk/products/cfd/overview`
[3] `http://www.x-plane.com/`

6.2 Interaction in Three-Dimensional Space

The techniques for interacting with three-dimensional virtual environments can be divided into three separate components: 1) selection and manipulation, 2) ravel, and 3) system control [100]. Travel is the process of moving within the virtual environment, selection and manipulation is how the user interacts with objects inside the virtual environment and system control is how the user interacts with the actual system hosting the virtual environment. This section now discusses each of these in turn with a specific focus on their potential to support the engineer engaged in three-dimensional engineering virtual environments.

6.2.1 Selection and Manipulation

Selection and manipulation covers how the user interacts with the virtual environment. Example techniques include gloves that attempt provide a sense of touch in response to the user 'touching' something in the virtual environment [104]; changes in air pressure created by ultrasonic waves [105]; vibration [106] and the multi-sensor, multi-buttoned controller supplied with the Vive virtual reality heat set from HTC[4] that implements a 'depth ray' that shines a beam of light from the controller into the virtual environment that allows the user to select and manipulate at a distance [107][100].

Moving away from state-of-the-art hardware and back towards the desktop PC, most users, including those at Airbus, interact using a keyboard and mouse and as such, there should be no introduction of new hardware and risk influencing any evaluation with the learning curve in using that new hardware. Literature also recommends providing feedback in response to the users actions [100]. This feedback can take the form of actually mimicking touch when the user clasps an object in the virtual environment to the physical 'click' of a mouse button. It is important that the system be responsive and show the visible reaction as instantaneously as possible to prevent the visible and physical feedback becoming disconnected.

6.2.2 Travel

Travel is the term used for how the user moves within a virtual environment. There are a number of metaphors used to describe the options available to the virtual environment with each metaphor existing in response to a particular type of virtual environment [108][100]. A flight simulator would obviously mimic flight whereas CAD and other

[4] `//www.vive.com/`. Last visited 2018-07-24.

engineering packages tend to use a targeted movement camera strategy[109]. This is one where the camera is targeted at a specific point in the virtual environment - typically the centre of the artefact (where the origin is set). Movement is achieved by ‘grabbing’ the scene/artefact and rotating the camera around the origin at a fixed distance, giving the impression that the artefact is spinning around its centre. This approach is typically uses in engineering virtual environments this stands to reason given these software packages are artefact centric.

One of the key challenges in designing a navigation strategy for any virtual environment is ensuring the user does not lose track of their location and so get lost. There are a number of strategies used in both engineering and other virtual environments that are applicable and solve this problem. The most appropriate strategy is highly dependent on the use-case however. Computer games will often use a ‘world-view’ map showing the user’s character and any other important information available to them. This becomes useful in an engineering virtual environment when exploring complex hierarchical artefact structures. *Are you viewing the near-side or off-side brake shoe?* for example.

ViewCube[110] from Autodesk Research is an example of innovation in the field of engineering virtual environment travel. It is a three-dimensional cube that allows the user to orientate the artefact while maintaining a visual key to the current orientation. It also allows switching to any other view by clicking on sections of the cube. While ViewCube structures and confines travel there is an argument that it does so at the cost of intuitiveness. Work along the lines of [111] discusses how tangible objects support design engineers. The argument against using ViewCube style travel control is that agency for the actual artefact movement is transferred from the user to the button and this interferes with the natural and intuitive behaviour of engineers interacting with objects/artefacts that is known to be important to engineering processes.

6.2.3 System Control

System control facilitates the user’s interactions with the application surrounding the virtual environment[100]. These actions can include opening/saving files, customisation and accessing help and guidance documents. In traditional two-dimensional Windows, Icons, Menus and Pointer (WIMP) interfaces[95][96], system interactions are typically accessed through menus situated at the edge of the screen/application window. Interactions are sorted and categorised to aid the intuitive finding and access. Given this, there is a strong argument to also implement menus for access to interactions like help and customisation.

A further aspect of the system control interactions that requires attention is how lists of documents/results are displayed and accessed. One of the most popular methods for displaying search results is to use a list [96]. The most common examples of these are two-dimensional user interfaces like internet search engines and while there it is clear their use is highly successful, translating this to three-dimensional interfaces is not straight forward. The examples in literature use pop-up menus within the three-dimensional virtual environment to display the results lists, however menu placement and image resolution mean users can struggle to interact with these efficiently [112]. Essentially the menu can scale such that text becomes too small or the menu itself could become obscured by the objects within the virtual environment. One solution is to use a pop-up 'tracking menu' [113] that is the technique used in [114], an engineering application where the menu takes the form of a 'navigation wheel' containing functions to control navigation. This provides the benefits of the ease of consistency in reading and interaction of a two-dimensional menu whilst maintaining the connection with the three-dimensional virtual environment.

6.3 Finding Information in Three-Dimensional Space

In the field of information search one of the most seminal works is the Information Foraging Theory [103] which relates user's behaviour to how animals seek food, following scents and restricting searching to specific diets. This work highlights the users natural behaviour when searching for information and should be reflected on when designing new users interfaces for search if the system is to not inhibit this natural behaviour and by consequence reduce intuitiveness and add to cognitive load. Information foraging is considered here because users will be searching for information/knowledge in the virtual environment of the model-based interface. This section now discusses Information Foraging Theory in more detail before exploring some information display techniques that could aid the finding of information and the discovery of new knowledge.

6.3.1 Information Foraging

Information Foraging Theory [103] identifies the similarity between the information seeking behaviour and the Optimal foraging Theory for food, or animal food seeking and feeding habits. The theories show the balance between the energy consumed in seeking food against the likelihood of finding the required information or food. Information foraging describes information seeking in terms of information and information patches, diets and scents. This section now discusses each of these in turn.

Information is the prey being searched for. It is the solution to an information need and so obtaining it is the goal of information seeking.

Information Patches are web pages, reports, photographs, etc. documents that contain information. Users will explore and evaluate patches against their ability to satisfy the information need.

Information Scents are the methods that users employ to evaluate Information Patches and decide whether each path is worthwhile exploring. Evaluation is based on a wide range of cues that provide insight into a particular patch and is evaluated against an information diet.

Information Diets are the attributes by which patches are evaluated. The diet is the wider features of the information need that provide context to the search to aid the filtering and selecting of information patches.

Within the context of an engineering virtual environment, a user will start a search with an information need, say the diameter of a bolt hole on a component. They will open and view a three-dimensional model of the component and see a range of information patches (component features). They will then systematically follow information scents based on the evaluation of patches against the information diet until the correct patch is identified and the information need is satisfied, or they will look at the component and examine a range of features until the find the right feature and from that, get the diameter.

6.3.2 Visual Representations to Aid Information Navigation and Knowledge Discovery

Information foraging describes how users behave when seeking information, the theory is not however an instructional manual on how to build effective three-dimensional virtual environments for information access and knowledge discovery. This subsection examines some visual interaction tools, techniques and strategies to aid the user on their journey to a successful information request and the discovery of new knowledge.

Visual Information Objects in GeoSpatial Information Systems

Geographical information systems such as Google Maps [115] represent geographic information in a manner that represents the features of the real world - a two-dimensional map. The benefit of geographical information system is that information is represented in a context that results in data being situated in the environment. The model-based approach to information navigation is similar in that it does the same in a three-dimensional manner. The approach can then be thought of as a GeoSpatial information system - a representation technique where information is associated to a particular position in space or to particular features/parts/subsystems/systems/regions. All governed by the product structure and its geometry.

Google Maps includes a points-of-interest marker or 'pin' (see Figure 6.1) that highlights specific and current locations of interest as well as allowing manual manipulation on the map itself. The user is able to click, drag, and zoom in and out. It is then appropriate to borrow these techniques and apply them to the virtual environment, with the user being able to manually 'move' around the model and zoom in and out to see detail. The concept of the point-of-interest pin can also be applied and **Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation** discusses the development of such a pin alongside a range of visual information objects: visual objects that represent individual and collections of documents (information patches) in the virtual environment.

Figure 6.1: An example of the Google Maps Point-Of-interest pin

Traversing the Product Structure

Engineering artefacts have an inherent structure both in terms of a component feature tree and in terms of the product structure. Figure 6.3 and Figure 6.2 are examples of the a feature and product tree as depicted in the FreeCAD CAD package. The feature tree is essentially how a component is made within the CAD package. The product structure is how components are related or the hierarchical structure of systems/subsystems/components.

The technique of traversing the product structure combines components into a single system or subsystem based on their position in the product structure hierarchy. The head, block, pistons and all other engine components being merged into a single file, i.e. the engine. To select the head for example, the user would first need to select the engine and on doing so the rest of the car would be replaced with all the components that make up the engine. Only then would the user be able to see and select the head itself.

The benefits of such an approach is that it leaves far fewer items on screen at one time when compared to viewing the entire artefact assemble in its entirety. This is problematic when visualising large, complex artefacts, even today's modern computers have a finite amount of memory and processing power and loading something such as an entire Airbus A380 is not feasible without high-end computing equipment. In theory then, this technique allows for highly complex artefacts to be visualised, assuming the product structure is hierarchical. The trade-off being the added effort required from the user in navigating to the desired component.

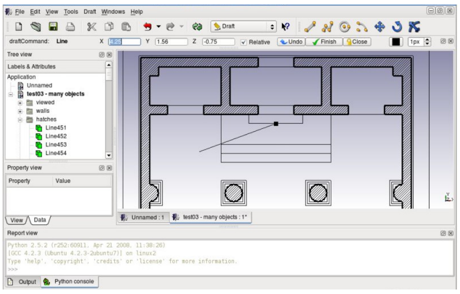


Figure 6.2: An example of a Feature Tree in FreeCAD

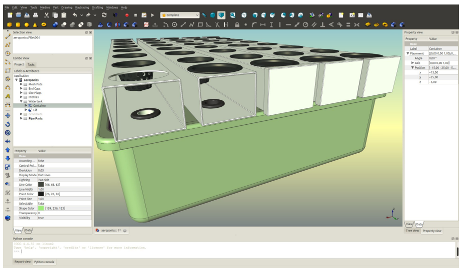


Figure 6.3: An example of the Product Structure in FreeCAD

Exploded Views

Exploded views are used within engineering to depict the relationships between components within an assembly drawing for example, by increasing the space between physically connected components such that all components become visible. See Figure 6.4 compared to Figure 6.5. The benefit of such an approach to information navigation is that components that would otherwise be obscured become visible and selectable while the spacial relationship between components is still visible. Both [116] and [117] show example of programmatically generating the exploded representation.

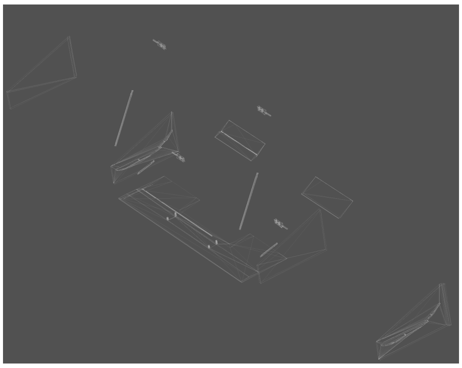


Figure 6.4: The exploded view of the front wing of a Formula Student racing car

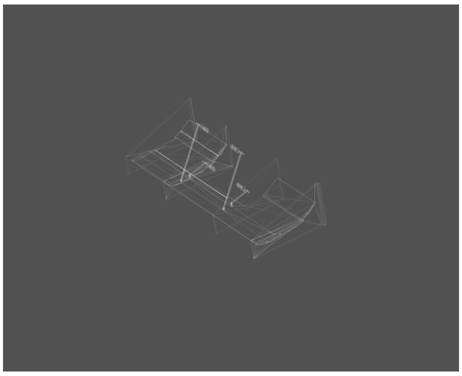


Figure 6.5: The unexploded view of the front wing of a Formula Student racing car

Cutaways/Sections

Cutaways are another engineering visualisation technique that remove a portion of the artefact to show internal features that would otherwise be occluded. The technique attempts to visualise the internal depths of an artefact while still depicting the overall physical structure. [118] presents an automated interactive technique that is capable of generating cutaways in complex intertwined three-dimensional artefact visualisations in response to the user selecting the cutaways dimensions. The authors examined a number of cutaway visualisations and determined that best practice for cuts should respect the geometry of any occluding parts as well as supporting interactive exploration. Both these findings are relevant to an artefact-based approach to information navigation as they attempt to maintain the user understanding of the artefact as a whole while supporting the finding of information.

[119] proposed a similar technique where tools are place in the virtual environment that allow the user to cut holes in the surface of an artefact and expose the inner parts. The implementation for the proposed system places a cylinder into the virtual environment that the user can select, orient and slide into the artefact model while viewing down the tube. Exposing the internal components to a depth controlled by the user.

Manual Manipulation of Artefact Parts

A simple method of navigating to the desired region of a complex artefact comprised of multiple components would be to allow direct manipulation of the component positions. The user could grab and move or click and hide components as desired. The draw-backs of such an approach is re-assembly and the user's ability to keep track of components once moved out of position. Particularly with large, complex assemblies. See Figures 6.6 and 6.10 for examples of components manually moved (Figure 6.6) and programmatically returned to the original position (Figure 6.7).

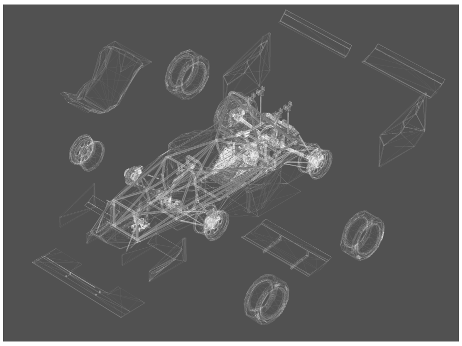


Figure 6.6: Formula Student racing car with a range of components moved

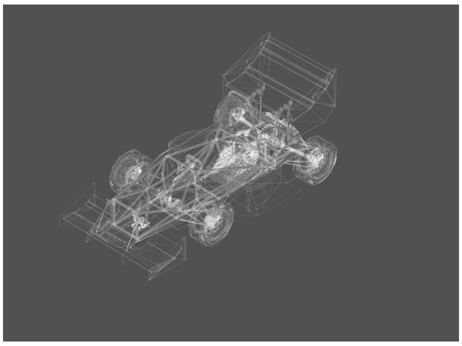


Figure 6.7: A complete Formula Student racing car

'MRI' Slicing

Medical MRI visualisations slice the artefact (in this case the human body) into a number of layered slices (see Figure 6.8). Visualisation is done by either presenting a single slice or a stack of slices. A version of this for artefact-based information navigation could allow the user to manipulate layers in the x, y, and z planes. When manipulated, this would have the effect of hiding sections of the artefact and allow the user to remove sections of the artefact that obscured the desired information object. This also has an added advantage over cutaways in that the user is in direct control of the portion of the artefact removed. See Figures 6.8 and 6.10 for examples.

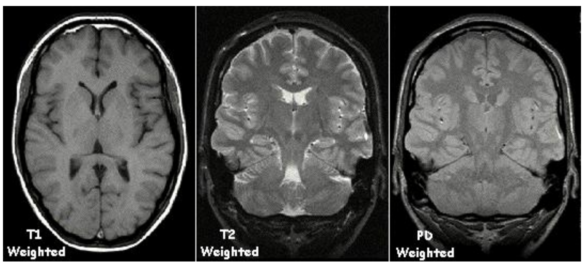


Figure 6.8: An example of MRI images moving through a human skull

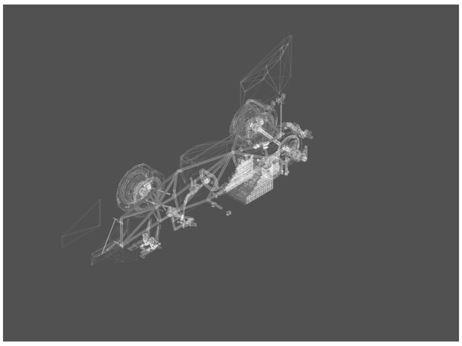


Figure 6.9: X-axis slicing of a Formula Student racing car

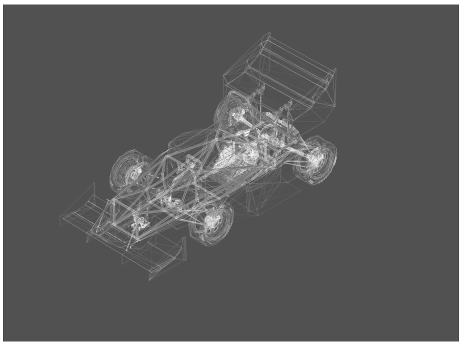


Figure 6.10: A complete Formula Student racing car

Textual Lists

Human computer interface and three-dimensional interface research also suggests providing the user with textual lists to aid navigation. This list provided alongside the proposed system would consist of the hierarchical product structure containing all features, parts, subsystems, etc. Selecting an item in the list would cause the user to automatically move to a position showing that particular item. [112] discusses menu placement within the three-dimensional environment however, problems with menu positioning and image resolution can result in the user either losing the menu or not being able to distinguish between menu items. Good two-dimensional design practice [120][121][122] places the menu at the side of the screen and in the foreground. See Figures 6.11 for example.

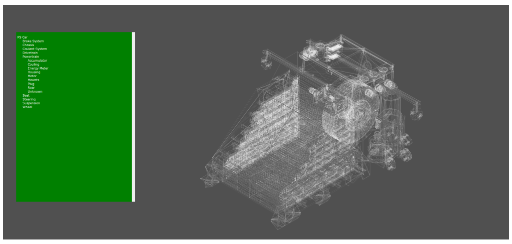


Figure 6.11: An example of a textual representation of a Formula Student racing car product structure with the power train components expanded

Heat Maps

In the booming field of information visualisation, heat maps are an efficient method for showing quantitative data [123]. In a heat map the data is coupled to the shade or intensity of a colour with the larger values typically tending towards a darker colour. Shading the parts of an artefact such that the darker shades represent the greatest volume of information lays a strong information scent that for example, can lead users to the most likely source of information. This opens the user to artefact-based knowledge discovery and not just information access. For example in the case of Airbus, use-cases might include the frequency of repairs, where document intensity displayed over the artefact provides a greater insight that the user could not obtain from a text based system alone without a degree of post-processing. See Figures 6.12 and 6.13 for examples.

Another method of using a heat map is alongside a traditional search text box to depict the importance of a query term to each component. As the users enters a query, each component is shaded based on its frequency or TF-IDF score (See **Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments**) from the documents associated to each component. An example of this in use could include searching for the areas of the car where a team member has worked. Searching for their name would return a heat map showing the level of input across the car.

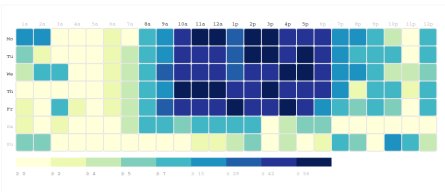


Figure 6.12: An example of a traditional Heat Map matrix plot

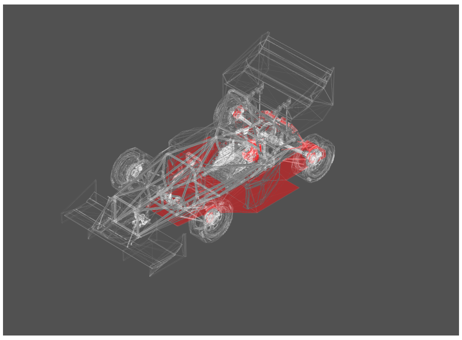


Figure 6.13: An example of a Heat Map depicted directly on the surface of a Formula Student racing car

6.3.3 Discussion

Information Foraging Theory found the similarities between animals foraging for food and the way in which people search for information. Typically, this is applied in text-based search engine environments. The model-based approach takes the user away from these text-based environments towards a visual and spacial interface. Arguably more akin to real-world foraging. In doing this, there is a risk that the challenges facing forages, namely seeing the food that's in front of them or buried within the leaves of a bush, are re-created in the model-based environment and users fail to find the information that is in front of them or buried within the depths of an engine (for example). The techniques outlined in this chapter aim to mitigate this.

The visualisation shows the user the full information landscape in a single representation and hence see where information is and is not situated. In comparison to a text-based system, this allows the user to evaluate and interrogate using strategies that are not available in cases where, for example, the text-based system returns too many or no results.

Research Question 2 aims to identify appropriate techniques for navigating the product structure and this chapter has derived a number of candidate techniques:

- Visual Information Object
- Traversing the Product Structure
- Exploded Views
- Cutaways/Sections
- Manual Manipulation of Artefact Parts
- 'MRI' Slicing
- Textual Lists
- Heat Maps

This list in not exhaustive but each technique is well-founded within its own field/literature and supports the task of finding information within three-dimensional virtual environments. As such it can be said that each approach is appropriate for the model-based approach of information navigation. As a means of some qualitative evaluation, the next section extracts some comments made by users from the study presented in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** and **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results** of this thesis.

6.4 Appropriate Techniques for Information Navigation within an Engineering Virtual Environment

As means of evaluating the list of navigation techniques and answer Research Question 2: *What are the most appropriate techniques for navigation information within* *a model-based virtual environment?* this section now evaluates a number of the comments relating to the navigation techniques implemented in the final study presented in **Chapter 8:Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** and **Chapter 9:Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**. Along with the findings from the actual implementation of the navigation techniques. The aim here being the generation of a validated list of navigation techniques for the navigation of information in engineering virtual environments.

The full details of the study are presented in **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**. As a brief overview, the study consisted of an A-B test between a text-based and model-based user interface to the same back-end search index. The study was run using Formula Student surrogate users, Formula Student data and CAD model, and a list of tasks based on the activities of the Wing In-Service team and translated to Formula Student equivalents. See Appendix A.3 for an overview of the 'user journey' for the system. Table 6.1 shows the list of tasks that participants were asked to complete. The description on how these were derived is shown in **Chapter 8:Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**. Tasks were written in pairs, with each participant asked to perform all 9 pairs of tasks. Participants were randomly assigned of either the text-based or model-based interface at the beginning of the session and the interface was then alternated for each pair of question such that, for example, task 1.a would be answered using the text-based interface and 1.b answered using the model-based interface (or randomly vice versa). The study was designed such that the dependent variable was the user interface (text-based vs. model-based). Table 6.2 shows this list of independent variables and how they were accounted for.

The study was run at four different universities, the University of West England (UWE), University of Bristol, University of Bath, and Imperial College London. Table 6.3 shows the breakdown of participants by university. In total, two hours were allocated for the study, this included 15 minutes at the beginning for a demonstration and 15 minutes for the participants to familiarise themselves with the system. As part of the study, participants were given a questionnaire based on the IBM Computer Usability Questionnaire [79] (see Appendix A.2) to complete at the end of the session. The comments presented in this section relate to those made by participants during the

Number	Question
1.a	Can you find the document '2018 Supplementary Rules' by the Institution of Mechanical Engineers?
1.b	Can you find the textbook called 'Advanced Brake Technology' by Bert Breuer and Uwe Dausend?
2.a	Do you think that past teams have completed sufficient research into tyres such that suitable tyres could be specified for the current car without additional research?
2.b	Do you think that past teams have completed enough research into impact attenuators that a past year's design could be re-used this year?
3.a	Name the area(s) of the car that have received the most computational fluid dynamics analysis.
3.b	Name the area(s) of the car that have received the most finite element analysis.
4.a	Name a past supplier of sprocket carriers.
4.b	Name a past supplier of electric motors.
4.c	Name a past supplier of brake pads.
5.a	Has anyone explored the use of additive manufacturing for the tripod housing?
5.b	Has anyone explored the use of additive manufacturing for the upright manufacture?
6.a	Find a front wing general assembly drawing. Did the author include their name and if so, who created the drawing?
6.b	Find the main hoop technical drawing. Did the author include their name and if so, who created the drawing?
7.a	Find a report on active aerodynamics. What is the file called?
7.b	Find a report on the use of analysis data to improve performance. What is the file called?
8.a	What area(s) of the car did [redacted] work on?
8.b	What area(s) of the car did [redacted] work on?
9.a	Who would you contact for advice about carbon fibre wheel rims?
9.b	Who would you contact for advice about front inboard suspension?
9.c	Who would you contact for advice about rear outboard suspension?

Table 6.1: A summary of the engineering tasks contained within 240 repair queries received by the Airbus Wing In-Service function during 2013. Names of individuals have been redacted for anonymity.

Independent Variable	Description
Effect of learning	Random order of questions
Indexing technique	Both interfaces use the same index.
Presentation of results	Both interfaces use a side-ways scroll and an image of the front cover as a preview. Accessing report is done via clicking on the image.
Bias in the tasks	Tasks presented in a random order and random interface assignment to each task (a or b).
Effect of browser/operating system	Chrome Web browser and Window 7 or 10 used throughout. No difference between performance on Windows 7 or 10.

Table 6.2: The list of independent variables and their description.

studies. The study was first run at the University of West England and then the user interface was updated to reflect the comments made.

Study Participants	
UWE	14
Bath	6
Bristol	19
Imperial	4

Table 6.3: A summary of the engineering tasks contained within 240 repair queries received by the Airbus Wing In-Service function during 2013

6.4.1 Findings from the System Implementation

The technology platform used was presented in **Chapter 4: Aim, Methodology and Research Questions**. The implementation of the system itself resulted in a number of insights into the methods of information navigation listed in Section 6.3.3 of this chapter that will now be discussed.

Firstly and most impactive was the realisation that the MRI approach was by far the easiest and simplest method of removing obstructing components and made the cutaway/section approach redundant. The cutaway/section approach involved placing a three-dimensional ‘tube’ into the model space and having the user orientate said tube into a desired angle (or three-dimensional vector) before moving the tube into the model, cutting away intersecting components. Relatively, this method required a lot of input from the user and, without any end-user testing, it was determined that the user could reach the same ‘physical’ position within the model using the MRI approach in a far shorter amount of time. The MRI approach was quicker, simpler to understand and far less cumbersome than manipulating a three-dimensional tube and as a result, the cutaway/section approach was removed.

The second realisation was a result of the technology on which the system was constructed. Components were loaded into the system as whole individual parts and for the MRI/cutaway techniques, where pieces of the model were removed, slicing individual component vertices was too processor heavy. This resulted in the system becoming unresponsive and ultimately unusable. The choice was then take to remove pieces of the model at an object level - if a component for example, is intersecting with a ‘slicing’ plane, then it is removed. Initially this was decided for implementation reasons, however, the results from **Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation** support this decision in that engineers show a preference towards interacting with artefact information at a component level.

6.4.2 Results

The results from the study presented here are qualitative and derive from either the comments from questionnaires, observations during the studies or from informal conversations with participants post study. A thematic analysis of the comments following the method outlined by Braun and Clarke [124] as presented by Moria and Bird [125]. The six steps of analysis: become familiar with data; generate initial codes; search for themes; review themes; define themes; and write up; were applied to user comments although themes were largely defined by the navigation and interaction techniques under investigation, or techniques that were not considered but were highlighted by participants, zoom for example. Given this and the narrow scope of themes, it was deemed appropriate not to formally define them.

The first point derived from the questionnaire was that users from the first study asked for the ability to zoom in and out of the visualisation. Certain components of the racing car were simply too small to view on the screen with a single, constant 'zoom level' and as such, users struggled to focus on particular components. In response to this, it was determined that the ability to zoom must be included in a set of appropriate techniques for information navigation within engineering virtual environments. Having added the function for the final three studies, the participants then noted there were some issues with its functionality, however, participants used the feature and as such its inclusion is justified. Table 6.4 shows a selection of comments relating to zoom functionality from both the UWE participants and from the later three studies. It is worth noting that once participants had zoom functionality, they then began mentioning the ability to pan, something that was built into the three.js visualisation software library and not considered before these mentions.

Comments Relating to the Heat Map feature
(UWE) '*Zoom required.*'
(UWE) '*a feature to zoom in would be useful*'
(UWE) '*I would appreciate a zoom function.*'
(UWE) '*scroll to zoom would be good*'
(UWE) '*Need to be able to zoom in and out.*'
'*Only tricky bit was zooming in where I wanted to as I would in CAD*'
'*the mesh of the car model made it tricky to see detail when zoomed in.*'
'*Zoom and pan had small issues*'
'*I think different zooming mechanisms would be easier.*'
'*the 3d model was difficult to zoom on individual components*'
'*Very intuitive, tiny detail , pan speed when zoomed in is too fast. That is my only problem, good job!*'

Table 6.4: A selection of comments relating to the requested zoom interaction feature

The second results was noticeable in observing the actions of participants during the study and were also reaffirmed during informal questioning post study. The textual list of components was initially constructed an indented list with all components visible. It was decided that an all-on-display approach would allow for the easiest finding of particular components. However, it was made clear that users spent a lot of time scrolling up and down the list in an attempt to find particular components. The length of the list was somewhat daunting. In response to this, a collapsible list was implemented. Table 6.5 shows the comments from before (UWE) and after the addition of a collapsible list functionality. It is interesting that one participant disliked the collapse feature after it was added.

Comments Relating to the Product Structure Text List
'*I don't like the dropdown boxes.*'
'*The model tree on the left of the car view would be must more useful if it was possible to maximise/minimise sections like with most filing system*'
'*3D system allowed traversing parts tree which was helpful for some searches.*'
'*The model tree on the left of the car view would be must more useful if it was possible to maximise/minimise sections like with most filing system*'

Table 6.5: A selection of comments relating to the product structure text list interaction feature

The third finding was explained by a number of users in informal post study questioning alongside one comment in the questionnaire (Table 6.6). Several users stated the system would be better if the controls were the same as the CAD package that they used. It seems that engineers have a preference for the methods of navigation that their particular brand of CAD package implements. Developing for this is not possible, other than implementing a highly customisable navigation method that allows users to mimic the navigation techniques of the CAD package in use. It is an argument for a longer term study of model-based information navigation however, this is not in the scope of this thesis given time constraints.

Comments Relating to Changing Controls to Match CAD
'*Maybe have the view controls adjustable to match SolidWorks or other CAD packages?*'

Table 6.6: Comments relating to changing the control such that they were the same as a CAD package

In addition to these changes, participants also commented on the heat map, onion peeling, drag, and reset mesh methods of navigating and interacting with the model environment. Tables 6.7, 6.8, 6.9 and 6.10 show the comments for the onion peeling, drag, and reset mesh respectfully. No comments were made of the exploded view methods although several participants were seen to use the feature during the study. One can see that for the most part, the comments are positive indicating the interaction methods were well received.

Comments Relating to the Heat Map feature
'*The heat map is great for finding who worked on what.*'
'*The search engine style was more intuitive, but the 3D model heat map displayed information well. If the selected part was highlighted on the 3D model that would be useful.*'
'*The visual aide and heat maps are useful when searching for items and understanding the overall layout of the car.*'
'*Love the drag function and heat map. The mesh reset was very helpful.*'
'*heat map can be misleading if documents talk about something but are not about that thing.*'

Table 6.7: A selection of comments relating to the heat map interaction feature

Comments Relating to MRI/Onion Peeling Feature
'*'onion' was a cool feature.*'
'*'Can be unclear what is toggled on (i.e. drag feature) and sliders aren't precise.*'

Table 6.8: Comments relating to the 'onion peeling' feature

Comments Relating to the Manual Manipulation of Components (Drag) Feature
'*Love the drag function and heat map. The mesh reset was very helpful.*'
'*Can be unclear what is toggled on (i.e. drag feature) and sliders aren't precise.*'
'*Rightclick drag function was way too sensitive.*'

Table 6.9: Comments relating to the drag feature

6.4.3 Discussion and Conclusion

In response the second research question, this chapter explored the fields of navigation in three-dimensional virtual environments from similar field and extracted those relevant to an engineering domain. Those features that were able to be integrated into the technology platform were then developed and tested though qualitative feedback from a range of Formula Student engineers. The results form the initial study generated some suggested improvements/new features (zoom) that were then implemented for the remaining studies. As a whole, the features in their entirety were well-received.

The output from this chapter and the answer to Research Question 2 is then a list of appropriate navigation and interaction techniques for the use in engineering model-based virtual environments:

Comments Relating to Mesh Reset
'*Love the drag function and heat map. The mesh reset was very helpful.* '

Table 6.10: Comments relating to changing the function allowing to reset mesh positions

- Heat Map
- Product tree List/Textual List
- Onion Peeling/MRI Slicing
- Drag/Manual Manipulation of Components
- Reset Mesh
- Exploded View
- Zoom
- Pan

The qualitative analysis used to generate these results does not fully validate the list and as such further research is required to generate a definitive list of techniques. For example, "Where's Wally?" style study, where participants are asked to navigate to a particular piece of information using the various navigation and interaction techniques listed. Using the time to find information as a measure, this would provide a quantitative evaluation and further verify the list.

Chapter 7

The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation

7.1 Introduction

Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments identified appropriate techniques for navigating within three-dimensional engineering virtual environment and in doing so briefly touched on the issue of the user being able to 'see' information sources. Information Foraging Theory shows the similarity between people searching for information and hunter gatherers searching for food. Using this analogy, being able to visually identify information sources is akin to seeing the fruit in a tree. Nature uses bright contrasting colours to visually separate fruits from the green leaves and maximise the likelihood of the hunter-gathers seeing and collecting the fruit (and ultimately spreading the fruit seeds).

Research Question 3: *What are the most appropriate techniques for displaying information within the model-based virtual environment?* sets out to identify the most appropriate techniques for displaying information within the three-dimensional engineering virtual environment. This is essentially attempting to recreate nature such that the user is able to quickly and easily identify information and ultimately maximise the likelihood of information dissemination.

The placement of some form of information markers within three-dimensional environments, whether virtual or not, requires careful consideration. Consider road traffic signs as an example, the UK government's Department for Transport publish an eight chapter manual [126] that cover every detail of the road signs that most drivers will pass without giving a second thought. The purpose of road traffic signs is to disseminate information with maximum ease and intuitiveness such that the driver receives and processes information while travelling at high speeds with relatively little cognitive effort. Consequently, the effectiveness of three-dimensional information visualisations such as road signs is an argument for a high benchmark in the visual representation of information in three-dimensional environments.

The UK Department for Transport achieve such a high standard by designing from the fundamentals of design theory and applying them to very specific use-cases. The work presented in this chapter outlines a similar journey with the aim of producing a number of strategies for generating Visual Information Objects (VIOs) within engineering virtual environments. VIOs being graphical markers or icons located in the VE that disseminate the existence and location of information within the three-dimensional space. This chapter begins with a more complete description of engineering virtual environments and the need to capture and disseminate engineering information. Then, some of the relevant fundamentals of design theory are covered. A set of use-cases are then distilled based on a range of discussions with industrial and academic engineers. The theory and use-cases are then combined and a number of VIOs are proposed before being evaluated through an end user study.

7.2 Background

This section outlines the background engineering VEs, three-dimensional computer generated visualisations and the current practices for the display of information within them.

7.2.1 Engineering Virtual Environments

The author of [127] and [128] defines virtual environments as interactive graphical displays that are sometimes enhanced with audio or haptic feedback. For the purposes of this study, engineering virtual environments are considered to be interactive graphical environments that support the engineer with engineering tasks. Engineering virtual environments are then product centric and allow for the design, drafting and testing of virtual artefacts. Computer Aided Design (CAD) systems being the most obvious example. Other examples include combining CAD with virtual reality [129], virtual assembly [130] and distributed virtual environments that allow engineers based at difference locations to collaborate via their own perspective virtual environments [131].

Augmented Reality (AR) systems mix virtual environments with the real world by overlaying computer generated images. Engineering examples include used AR for manufacturing planning, for training and architectural construction and inspection and renovation [132][133][134]. A sibling research field to VE, there are many elements of AR research, such as information visualisation and the design of labels, that can be directly applicable to virtual environments.

CAD packages cater for both two-dimensional and three-dimensional drawing with standards extending from traditional pencil and paper drafting. Two dimensional drawings rely heavily on a few basic visual building blocks (predominantly line and text) and use them in a very precise and controlled manor [135]. Figures 7.1 and 7.2 are examples of how lines, arrows and text are used to represent additional information along side the visual representations.

The combination of lines and text form VIOs that appear highly efficient in their dissemination of engineering information. In the wider VE/data visualisation fields these VIOs are categorised as external labels. Labels in virtual environments are categorised as either internal or external. Internal or surface labels are those more associated with maps and are written on the surface of visual objects or 'spatially bound' to an object [136]. External labels are more associated with engineering drawings [135] with labels being placed surrounding the visual object and coupled with lines to convey the connection between label and object. These labels are also standardised across three-dimensional drawings through the ASME Y14.41 and ISO 16792 international standards.

Based on the prevalence of external label use in engineering one could argue for their wider use outside of the CAD package. However, their use in non-engineering virtual environments and AR have shown their use causes the VE to become cluttered

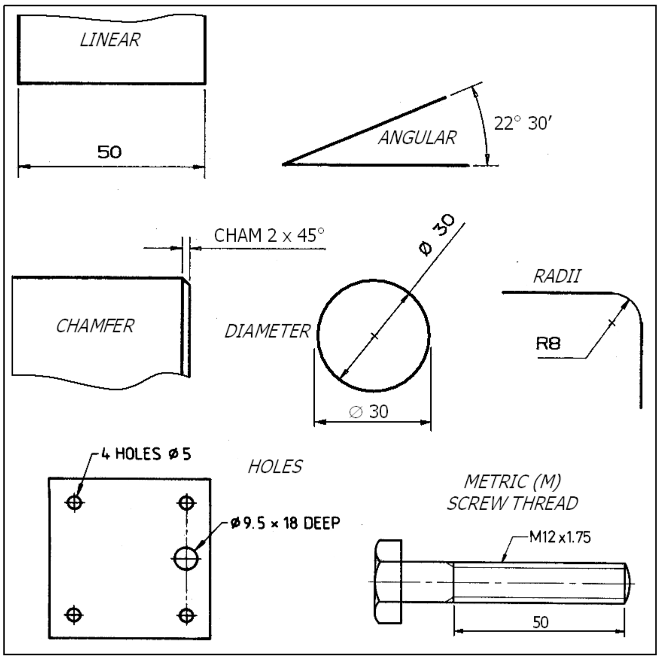


Figure 7.1: Examples of VIOs in engineering drawings

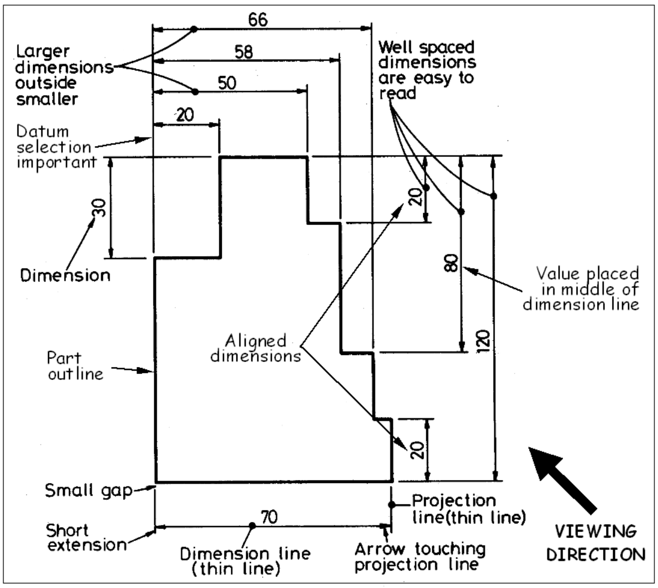


Figure 7.2: Examples of VIOs in engineering drawings ant their descriptions

and issues with occlusion arise. To overcome this, view management techniques are implemented to manipulate the positioning of labels and connecting lines based on the orientation of the user [137][138]. The alternative is to separate the text from the VIO and use an icon to identify position, similar to those use in Geo-Information Systems (GIS), the Google Maps pin for example [115]. Icons, when well-designed, can aid the efficient dissemination of information [139].

7.2.2 Engineering Information Management Across Product Life-Cycles

[140] discusses data, information, knowledge and their relations within the context of engineering design, outlining their importance before describing a formal framework. Engineering information management (EIM) then aids the efficient capture and dissemination of information in support of business processes, whether they be software application or business processes/practices themselves [141].

Engineering information usage can range from supplier information to records of previous designs [142] recorded in the forms of written notes, sketches and printed and annotated CAD drawings to name a few examples [143]. A whole life-cycle approach to engineering and engineering information can improve the engineering process, for example, a product's end of life condition data being fed back into the design and manufacture to produce a more robust evolution of the design [144]. It is here where the likes of product life-cycle management and product data management systems aim to facilitate the capture and reuse of information to maximise data reuse, knowledge discovery and consequentially the product itself.

One of the key features of EIM, whether handled by a dedicated PLM/PDM system or not, is the finding and re-finding of information via a search engine [50] and while current enterprise search applications can struggle to match those of the internet [145], work is progressing to better understanding the challenge of enterprise search and its improvement [41][90][146]. [74] outlines a visual strategy for such an improvement, based on the premise that engineers think visually and functionally, the authors describe a strategy for the navigation of engineering information via a three-dimensional artefact. An important area that is not considered by [74] is that of user experience/interface design and in particular the best or most appropriate way to visually represent information given the user and use-case. This is the challenge addressed by this chapter.

7.2.3 Engineers and Complex Engineering Artefacts

Websites such as https://www.thingiverse.com/ and https://grabcad.com/ are representative of the vast ranging varieties of artefacts drawn in CAD systems. Artefacts vary in shape from circular to square, naturally inspired to highly engineered and from simple single components to massive multi-component and multi-system assemblies. As part of an exploration of the world of PLM, [147] examines the typical number of parts in various products, listing a can of deodorant having 20 parts, a car having 25,000 parts and an aircraft having 40,000 parts. From these it can be said that the engineer can then be faced with any number, size, shape and complexity of artefact when interacting with engineering VE. Designing VIOs that are distinct from the artefact VE and instinctively seen by the user with minimal cognitive load is not a trivial task.

7.3 User Considerations

The role of the user in Human Computer Interaction (HCI) is a maturing field with a vast range of best practices and considerations documented [120][122][148]. This section highlights some of the most applicable user considerations for engineers and engineering virtual environments that apply to the design of engineering VIOs.

7.3.1 Memory Load

Memory load is the cognitive effect required to interact with the display. The user should not have to remember information between actions. One method of achieving this is to design tasks and actions that only take a few steps to complete. The environment should be designed for both novice and experienced users. A novice user will require labels providing more detail while experienced users should be provided with short cuts [148][149].

7.3.2 Information Assimilation

Information assimilation should be easy for the user and a good environment should facilitate this. The environment should assist rather than inhibit the user's ability to view and understand the information being displayed. Techniques for achieving this includes formatting the display in a fashion that is familiar to the user and in a way that is suitable to the task, tables of columns and rows for numerical data for example [148][149].

7.3.3 Information Foraging

Information Foraging highlights the behaviour similarities between users searching for information and animals foraging for food [103]. Users are said to follow information scents that steer the path of the search towards an information goal and switch between scents when a particular path is proven fruitless. The manner by which Information Foraging works within a VE is highly dependent on the nature of the VE. The system outlined in [74] for example highly constricts the 'scent' to the product structure.

7.3.4 User Control

User control makes the user a participant in the visualisation. [122][148] both advocate the user being given control of various aspects of the display with [148] stating the importance of maintaining the users sense of locus control in dialogue design. [122] states that control is achieved when a person, working at his or her own pace, is able to determine what to do, how to do it and is then easily able to get it done.

7.4 Design Considerations

One of the most significant challenges addressed in this chapter is creating VIOs that are distinct from the VE itself. AR implementations achieve this with relative ease given a computer generated VIO graphic will clearly stand out against a real world background. Some AR research focuses on the relative positioning of VIOs, occlusion and readability of text [138][150]. Each of these is highly applicable to VIOs in virtual environments and so any solution must adhere to the best practices of each of these (as well as design in general). The following sections set out the primary design considerations, with example and best practices highlighted throughout.

7.4.1 Colour

Colour is one of the most power yet often misused building blocks of visualisation [120]. It can be soothing to the eye, liven a dull visualisation, encode discrete values, reinforce logical organisations, draw attention, evoke emotional responses and highlight/create relationships [120][148] but when used incorrectly it can distract or insufficiently highlight important information [120]. To counter this, [148] advocates designing in monochrome first and then adding colour and, in line with [120] and [122], repeatedly highlights that colour should be used sparingly, thoughtfully and simply. For the purpose of highlighting information within engineering virtual environments, a VIO within the VE must be coloured such that it appears distinct from the VE.

7.4.2 Shape

Shape is one of the core building blocks of information visualisation. It is useful for labelling and encoding categories but can have significant cultural connotations (cross for example) that should be taken into consideration [148].

When examining shape in visualisations it is useful to de-construct some of our more common information visualisations. The bars on a bar chart are rectangles whose lengths reflect quantitative data and a pie chart is a circle that represents the whole of something that is divided up into parts. Shapes can be bound directly to data in this way to allow for the direct visualisation of information [151].

Within engineering virtual environments using shape to form distinct VIOs is not a straight forward solution. Given a CAD engineer can be designing many free format regular shaped parts, there is no single means in generating a VIO that is consistent in shape and distinct to the VE.

7.4.3 Size

Size can be perceived as the importance or magnitude of an entity/value. The larger an entity the more important the entity or the larger the underlying data. The human mind is good at comparing the area of a rectangle when the length changes (as in bar graphs) but struggles when both length and width are varied at the same time. In the same vein, users struggle to comprehend the differences in the areas of circles although circles are good at providing coarse representations [123].

Within the context of virtual environments, size suffers from the same problem as shape in that there is no way of determining in advance the relative sizes of parts particularly in large sub-assemblies, hence there is no guarantee of a consistent relative size within the VE. It is possible to always have a VIO that is larger than the artefact but whether this is practical is another matter.

7.4.4 Icon

Icons are more complex forms of shape [148] or to put it another way, icons consist of simple symbols that themselves consist of graphical elements [139]. Well-designed icons can help users act quickly and surely, represent visual and spacial concepts, save space in the visual display, speed up search, are immediately recognised, provide better recall and do not rely on the user's ability to read [139].

Icons still suffer from the problem of consistent differentiation from the engineering VE however, a reasonably recognisable icon can move to counter this. For the reasons outlined, considerations of size, shape and icons alone is insufficient to generate the required visual distinctiveness.

7.4.5 Line

Lines have a number of practical uses in information visualisation depending on how they are configured and positioned [148]. The weight of a line can represent magnitude or importance but users can struggle to discriminate between small differences in line thickness. Line endings can be modified with dots, forks or arrow heads to encode different functions (source, destination, relationships, etc.). Lines can be drawn as patterns (solid, dashed, dotted, etc.) and have specific meaning, as previously discussed, on engineering drawings for example [135].

7.4.6 Text and Typography

Text and typography should be used sparingly as it can clutter a visualisation and can heavily influence the users gaze when trying to interpret the visualisation [123]. In terms of font, Serif fonts are deemed better for blocks of texts and Sans-Serif for titles, tags and labels. [123] also asks designers to be aware that writing in all-caps can take longer for the mind to process and adds unnecessary cognitive load for the user.

In three-dimensional VE text tends to fall into two categories, internal and external labels. Internal labels or surface labels are most often associated with maps and are written on the surface of visual objects or 'spatially bound' to an object [136]. They are frequently seen in the field of biology where the names of parts of the body are written directly on the surface of the image [152]. External labels are more commonly associated with engineering drawings [135] with labels being placed surrounding the visual object and coupled with a connecting line to convey the relationship between label and object.

[150] performed an analysis on the readability of text in AR and recommended the use of 'billboarding' text by placing it on a solid opaque background coloured to contrast with the colour of the text to help distinguish the text from the varying background. This technique may also help in engineering virtual environments. However, the work presented in [137] and [138] shows how labels in three dimensions can quickly clutter the

VE and become difficult to read without a good view management system in operation.

7.4.7 Figure/Ground (Gestalt Principle)

Figure/Ground is a Gestalt principle and states that the mind separates the visual field into figure (the foreground) and ground (the background). Graphical user interfaces use figure/ground with great effectiveness in the design of most desktop environments and applications. The placement of menus and other system features around the edges of the screen that open up over the focus of the screen separate the two. CAD systems use figure/ground in the same manner, with tool sets and systems menus all kept in the foreground and the artefact space/VE pushed to the background.

7.4.8 Consistency (Gestalt Principle)

One of Gestalt's Principles of graphic design is consistency which when applied to virtual environments can be considered at two levels. In the first, consistency is key, regardless of the exact aspect of the display in question. Whether that be label and display elements (headers, footers, menus), dialogue, data visualisations, actions, functions, operations and even the positions of functional elements. Inconsistencies increase the cognitive load on the user and unnecessarily complicate the interaction and dissemination processes.

On the second level, inconsistency can be useful. In data visualisation inconsistencies can reflect and draw attention to inconsistencies in the data. With information visualisation within virtual environments, an inconsistency between the colour/size/shape of the VIO and the VE itself can emphasise the separation between the two visual elements and reduce the memory load.

7.5 Engineering Information Markers in Three-Dimensional Space

Between 2015 and 2017 regular meetings and discussions took place between the author of this thesis and engineers, project managers and engineering academics with the aim of determine the types of VIOs required within engineering virtual environments. These interactions were based around test bed systems (see **Chapter 4: Aim, Methodology and Research Questions** and [74]), knowledge of existing CAD systems and 'real world' industrial use-cases and focused on the types of information based interactions that engineers have with three-dimensional product representations. Example

Information Object	Description	Example Usage
Component	Single/Multiple product components.	The remoulding of the driver seat to a new driver.
System/ Subsystem	A collection of components that perform a specific task.	The design and testing of the fuel system.
Point	A specific x, y, and z co-ordinate within the product co-ordinate system.	The location of a hole in the exhaust system.
Vector	An indicator of movement or direction.	Adjust the pedal position towards the forward of the car to suit the driver's leg length.
Layer	The geometric layers of a CAD model.	Removing the body work to inspect the chassis.
Feature	Holes, chamfers, rounds, fillets, etc.	Examine the bolt hole positions for assembly/disassembly.
Section	A specific area covering of one or more components.	The aerodynamics testing of the front of the car.
Region	A specific area that is not component dependant.	Corrosion along the training edges of the front spoiler components.
Surface	A single or collection of component surfaces.	Inspecting the quality of the paint.

Table 7.1: Identified Infromation Objects

of these interactions include searching for product related reports, investigating clash detection/interference and general product related communication.

Table 7.1 shows the information objects/interactions identified. Each information object is related to a geospatial property of the physical product and is shown alongside a description and example usage. The goal then is to design a range of VIO that encompass each of these information objects/interactions while adhering to good design practice and minimising the user effort required to find information.

7.6 Visual Information Objects

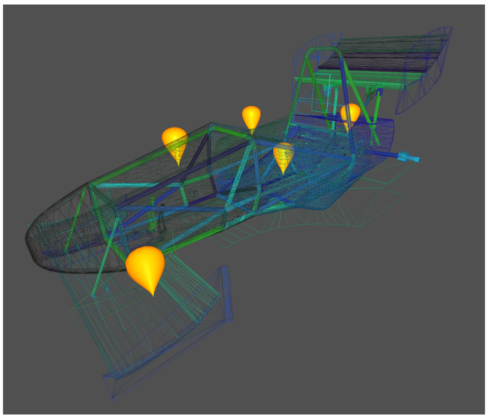
Based on the interactions identified in Section 7.5 this thesis proposed four types of VIO: Point-Of-Interest (POI), Component-Of-Interest (COI), Region-Of-Interest (ROI), Section-Of-Interest (SOI). Table 7.2 shows each of these VIOs and the corresponding interaction covered.

7.6.1 Point-Of-Interest (POI)

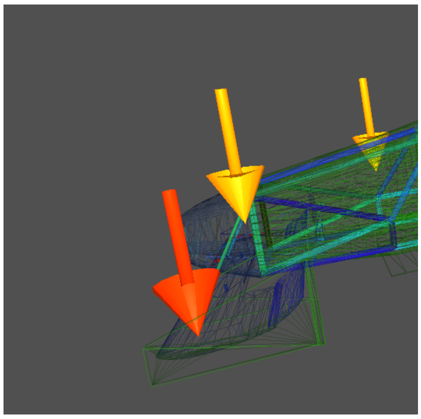
A point-of-interest represents a piece on information that relates to a single location or direction, work was carried out at position x, y, and z or move the seat forward for examples. Example of POIs exist in both literature and current market products, think of the Google Maps pin for example [115]. The version shown in Figure 7.3a show

Proposed VIO	Information Object
Point-Of-Interest (POI)	Point
Point-Of-Interest (POI)(Directional)	Vector
Component-Of-Interest (COI)	Component System/Subsystem
Region-Of-Interest (ROI)	Layer Region Surface Feature
Section-Of-Interest (SOI)	Section

Table 7.2: Proposed VIO and corresponding interaction covered



(a) POI



(b) POI (Directional)

Figure 7.3: POI (Directional and Non-Directional)

orange three-dimensional non-directional markers and Figure 7.3b shows the directional version.

7.6.2 Component-Of-Interest (COI)

Components-of-interest were highlighted as a possible improvement on POIs in [74] where a proof of concept using POIs was preformed. COIs cover the component and system/subsystem techniques for CAD interaction. Figure 7.4 shows the car seat coloured pink/purple.

7.6.3 Region-Of-Interest (ROI)

Region-of-interest are a direct requirement from industry following a discussion on the need to be able to highlight a particular section on the surface of an artefact, indicating corrosion across a section of a single or multiple components for example. ROIs cover layers, regions, surfaces and features. Figure 7.5a show markers that would represent

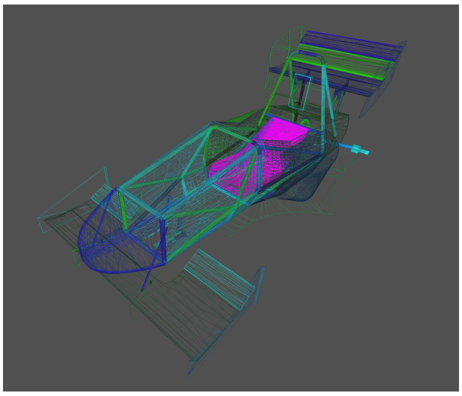
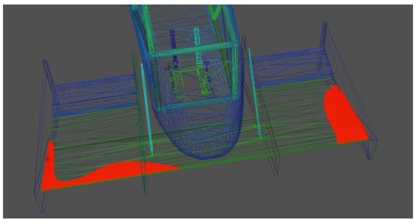
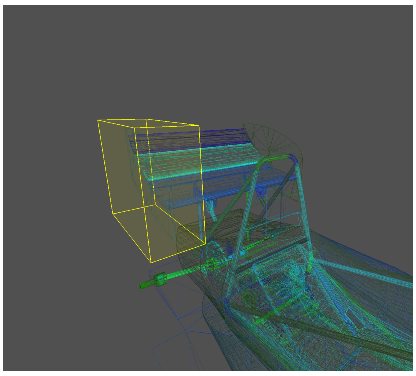


Figure 7.4: COI



(a) ROI



(b) SOI

corrosion on the fornt spoiler of the car.

7.6.4 Section Of Interest (SOI)

The concept of Sections of Interest were viewed in an industrial use-case where engineers drag two and three-dimensional wire-frame boxes over regions of the product to select a range of components. Given the model is depicted using wire-frame, Figure 7.5b shows how the faces of the SOI are shaded but transparent to maintain visibility of the underlying model and still be primarily a three-dimensional wire-frame box.

7.6.5 Information Visualisation Design Theory in Practice

While there are numerous design theory elements, Table 7.3 shows a list of those with deliberate implementation. The goal throughout is to make the identification and assimilation of VIO instinctive and with minimal cognitive load for the user.

Theory	Implementation
Colour	• The darker shade background causes the colours of the model to appear brighter and so clearer. • Colour pallet for the model is limited to blue/green keeping reds for VIOs to maximise the contrast and develop consistency. Also,warnings and danger tend to be reds, orange and yellows and so the user should be instinctively drawn to the VIOs [123][148].
Shape	• The model is drawn in wireframe allowing for high level of transparency. • The VIOs are solid maximising the visual differences.
Icon	• The POI VIO is based on the widely used Geo-Information System 'pin', Google Maps marker [115] being one of the most common. This familiarity will lower the cognitive load required to interpret the VIOs purpose. • The cluster marker is clear to read and simple and aimed to intuitively represent and inform of the number of clustered markers. Orange coloured outlines are consistent with the VIO colour scheme and the white background and black lettering allow for ease of reading.
Size	• Markers are always clear and visible with the LOD strategy insuring so. • Cluster VIO icon reacts to the distance the user is from the model such that it is always readable.
Figure/ Ground	• Markers and model are all placed in the foreground making them the main focus of the environment.

Table 7.3: Implemented Design Theory

7.7 Study

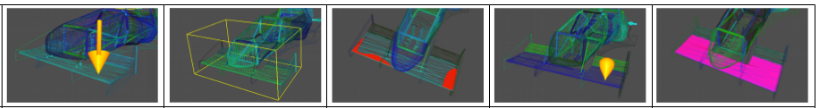
In [75] the evaluation of human-computer interfaces for information retrieval is discussed and includes the importance of participants being the intended end users of the system. In line with this, a group of 32 engineers from both academia and industry participated in the study. The least qualified participants were either third year students and a current member of the Formula Student Bristol Electric Racing Team (two participants) or masters students (five participants). The remaining participants were either academics ranging from PhD to professorial level (14 participants) or currently working in the engineering industry (11 participants).

In addition to the information objects derived from the discussion mentioned in Section 7.5, a range of seven information seeking scenarios were also determined (see Table 7.4). These were largely based on the practices of the Wing In-Service team based in Filton (discussed in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**) as engineers responded to maintenance and repair requests. These scenarios were used to evaluate the VIOs.

Participants were given a questionnaire showing screen shots of the five markers alongside the information seeking scenarios (See Appendix A.1). See Figure 7.6 for an image of the questionnaire used. Screen shots were used to ensure users were focused on

Looking at the screen shots below, rank the markers from most (1) to least (5) appropriate base on whether you would use the marker for the association of the information source with the model.

Name:



Competition guidelines regarding a part/system/sub-system					
Details of the structural testing for a part/system/sub-system					
Damage report for a hole in a part/system/sub-system					
Technical details for a part/system/sub-system alignment adjustment					
Aerodynamics report for a part/system/sub-system					
Damage report showing locations effected by corrosion on a particular part/system/sub-system					
Livery and paint specifications for specific part/system/sub-system					

Figure 7.6: An image of the questionnaire used to evaluate the VIOs

the markers themselves rather than other aspects of the system such as the navigation of the artefact space or system performance. Users were asked to rank the VIOs in the order which they would use the marker to represent each particular information source.

Information Seeking Scenario
Competition guidelines regarding a part/system/sub-system
Details of the structural testing for a part/system/sub-system
Damage report for a hole in a part/system/sub-system
Technical details for a part/system/sub-system alignment adjustment
Aerodynamics report for a part/system/sub-system
Damage report showing locations effected by corrosion on a particular part/system/sub-system
Livery and paint specifications for specific part/system/sub-system

Table 7.4: The list of information seeking scenarios generated through discussions with Airbus personnel. See **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** for more details.

In summary, Table 7.5 shows the list of participants in the study, Table 7.4 shows the list of scenarios that participants were asked to consider and Figure 7.6 shows the questionnaire that was used to present these scenarios and the five VIOs. Participants were informally asked to complete the questionnaire in an ad-hoc fashion and not in a formal study setting, i.e. during a meeting with their personal tutor. No additional information other than that contained within the questionnaire and participants were given as much time as they required to complete the questionnaire.

Participants	Number
Formula Student Bristol Electric Racing Team (third year undergraduate)	2
Formula Student Bristol Electric Racing Team (Masters)	5
Academic Engineers (PhD to Professor level)	14
Industrial Engineers	11
Total	32

Table 7.5: Implemented Design Theory

7.8 Results

Table 4 shows the VIO's ranked first and second for each of the information seeking scenarios. For thoroughness the results were processed using three methods. The first using the percentage ranking, i.e. the number of people who ranked each marker first and second against the total number of participants. The second approach applied a score of 5 to 1 to each of the ranking positions ((position: score): 1: 5, 2: 4, 3: 3, 4: 2, 5: 1) and the total score for first and then second were summed. The third approach was derived from participant feedback where participants stated they sometimes struggled to differentiate between the ranking of two different markers and would score them equally if the questionnaire allowed. This approach sums of the scores for first and second position and calculates the percentage against the total number of participants. Essentially this takes the mid-ranking point (third) and working on the premise that the two ranked above third position are preferred over the two ranked beneath third position.

The results show that overwhelmingly the participants preferred the use of COI's for displaying the information sources with each of the other VIO's being used occasionally for specific purposes. Across the three methods of analysis, there was very little change in the highest two VIO rankings for each of the information sources with the results staying stable regardless of the method used and no difference at all between the first two methods. The main differences in the final analysis occur due to the approach preventing the same VIO appearing in both first and second position. This benefits the SOI marker in two of three cases with the ROI ranked second in the third case.

There is no definitive conclusion on the use of directional and non-directional POI markers. There appears to be a general preference for the directional version. On the occasion where they were ranked first and second, the directional version ranked higher. However, the information source in this case was the position of damage (a

Information Seeking Scenario	Position	VIO					
		Percentage		Score		Total 1st + 2nd (%)	
Competition guidelines regarding a part/system/sub-system	1st	SOI	50	SOI	80	SOI	69
	2nd	COI	38	COI	48	COI	62
Details of the structural testing for a part/system/sub-system	1st	ROI	41	ROI	65	ROI	47
	2nd	COI	31	COI	40	COI	53
Damage report for a hole in a part/system/sub-system	1st	POI(D)	50	POI(D)	80	POI(D)	72
	2nd	POI	47	POI	60	POI	66
Technical details for a part/system/sub-system alignment adjustment	1st	COI	28	COI	45	COI	56
	2nd	COI	28	COI	36	SOI	24
Aerodynamics report for a part/system/sub-system	1st	COI	53	COI	85	COI	88
	2nd	COI	34	COI	44	ROI	53
Damage report showing locations effected by corrosion on a particular part/system/sub-system	1st	ROI	88	ROI	140	ROI	88
	2nd	COI	53	COI	68	COI	53
Livery and paint specifications for specific part/system/sub-system	1st	COI	63	COI	100	COI	91
	2nd	COI, ROI	28	COI, ROI	36	SOI	50

Table 7.6: Information sources and the corresponding highest two ranked VIOs

hole) which means the directional marker was preferred for an information source that does not contain directional information. For the information source containing the directional information the COI was preferred followed by the directional POI. There is then a question of whether the POI markers warrant further study although is highly reasonable to say that their intended use could be learned during the use of an artefact-based information navigation system. There is also an argument to remove the concept of the directional POI given participants preferred the COI for the directional information. The directional POI marker design could then be used as a non-directional marker, following how the participant used the markers. At the very least, a note should be made of the possible slight ambiguity around their general use.

Information	VIO
Rules/Guidelines	SOI, COI
Structural	ROI, COI
Technical Adjustments	POI(D), POI
Surface (aerodynamics, livery)	COI, ROI, SOI
Damage	ROI, COI

Table 7.7: Usage guidelines for information type and VIO

7.9 Conclusion

This chapter presented the design and verification for a number of Visual Information Objects (VIOs), visual markers that indicate the presence of information within a three-dimensional engineering virtual environment. Five marker designs were developed based on input from 10 Airbus engineers and project managers and a range of user and design considerations. The resultant VIOs were then verified using 36 engineers from both academia and industry using a number of information seeking scenarios.

The overwhelming conclusion from the results is that engineers prefer to use Component-Of-Interest (COI) markers (based of the artefact components themselves) for the displaying information. This ties with the understanding that engineers think visually and functionally in-line with the artefacts that they design and build. The COI markers are designed to mimic the real world artefact and so this preference is expected. There are however times when each of the other VIO's are preferred, demonstrating both their need for a range of markers and verifying the design process which lead to their development.

A limitation in these results however was the choice of VIO colour used in the study questionnaire. While effort was made to limit the colour palette of the racing car to blue/green, the background to grey, and the red/yellow for the VIOs, the differing in colours between the VIOs could have an effect on the results. For example, the bright pink COI could have drawn the eye more that the yellow SOI and swayed the participants towards COI over SOI. Further study is also required to eliminate this limitation.

Taking these limitations into account, Table 7.7 shows the suggested usage guidelines for the VIOs based on the results presented in this section. When displaying Rules or Guidelines, for example, one should use either a Section-Of-Interest (SOI) or a Component-Of-Interest (COI). These information types listed are those taken from the scenarios from the questionnaire (Figure 7.6).

7.10 Discussion - Layer Of Detail (LOD)

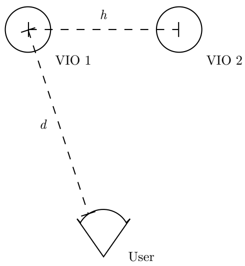
The work presented in this section, a Layer Of Detail (LOD) strategy, was not evaluated as part of the study presented in the previous section and as such is presented separately here. The Visual Information Object (VIO) designs developed in this chapter are only one step in developing a system that enables engineers to navigate information via the product structure. Another aspect that is closely related to the VIO design is a Layer-Of-Detail (LOD) strategy. As the number of information sources increases, the virtual environment is at risk of becoming visually cluttered. This would make the system unusable and, as the user zooms out from the artefact, small markers could shrink to the point where they become too small to be visible. An LOD strategy manages these by, for example, clustering markers into one more visible marker that identifies itself as a cluster of other markers.

GIS like Google Maps [115] use various layers of detail implement zoom features that allow the user to view a map at varying levels of detail: from the entire planet earth to continent, country, city, and road for example. Loading the detail of every road and building when the user is viewing an entire planet is impractical given the volume of information and the computational power of most computers. The LOD approach here loads different visualisations with varying amounts of detail based on the user's zoom level.

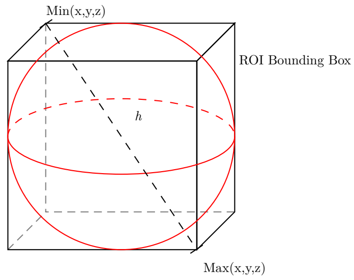
CAD files can become very large in the physical memory size and in response level of detail tools within CAD packages were designed to produce lower detailed versions for export and sharing. Details such as bolt holes and smaller component such as nuts and bolts can be omitted from the representation without losing the overall essence of the artefact. It is envisioned that these tools/strategies would be used to generate artefact models of varying detail that are then loaded based on the level of zoom. However, for the purposes of the test bed system presented here, exporting the FS model as a low resolution STL and compressing using OpenCTM has resulted in the test bed system not requiring an LOD strategy for the model but such a strategy will be needed for implementations with models larger in size.

The second use is for the reduction of clutter in visualisations of high VOI density. Examples of this can be seen in Google Maps marker clustering [115] where VIOs are merged and the cluster identified using a different shape and the number of VIO contained within.

Figure 7.8 shows such a strategy in action and the user zooms away from the car



(a)



(b)

Figure 7.7: Determining the point to cluster for both separate VIOs (a) and the specific height component for individual RIOs (b)

model. The system measures the distance between each VIO and camera and clusters if the ratio of screen distance falls below a configurable value. See Figure 7.7 and Equations 7.1. The cluster VIO is place equidistant between the two original markers.

The final aspect of the LOD strategy relates specifically to ROI. Given the unpredictable nature of their shape, these VIOs risk becoming too small to be visible at varying levels of zoom. The solution to this involves the system calculating the distance between the maximum and minimum vertices of box bounding each ROI (see Equation 7.2). If this distance falls beneath the configurable value used for the cluster size calculation, the ROI transforms into a POI while maintaining the ROI colour. This marker is placed at the calculated centre of mass of the ROI. See Figures 7.9a and 7.9b.

$$\frac{h}{d} \cdot \frac{screenWidth}{screenHeight} < clusterSize \tag{7.1}$$

$$\frac{|maxVertex - minVertex|}{d} \cdot \frac{screenWidth}{screenHeight} < clusterSize \tag{7.2}$$

7.11 Discussion and Conclusion

This chapter set out to answer Research Question 3: *What are the most appropriate techniques for displaying information within the model-based virtual environment?* Through a range of discussion with industrial and academic project managers and engineers, and a review of design theory, a range of Visual Information Objects (VIO) were designed and tested. The results of which showed that engineers have an overwhelming preference for navigating information via product components or Components-Of-Interest with the other forms of VIO preferred in particular cases.

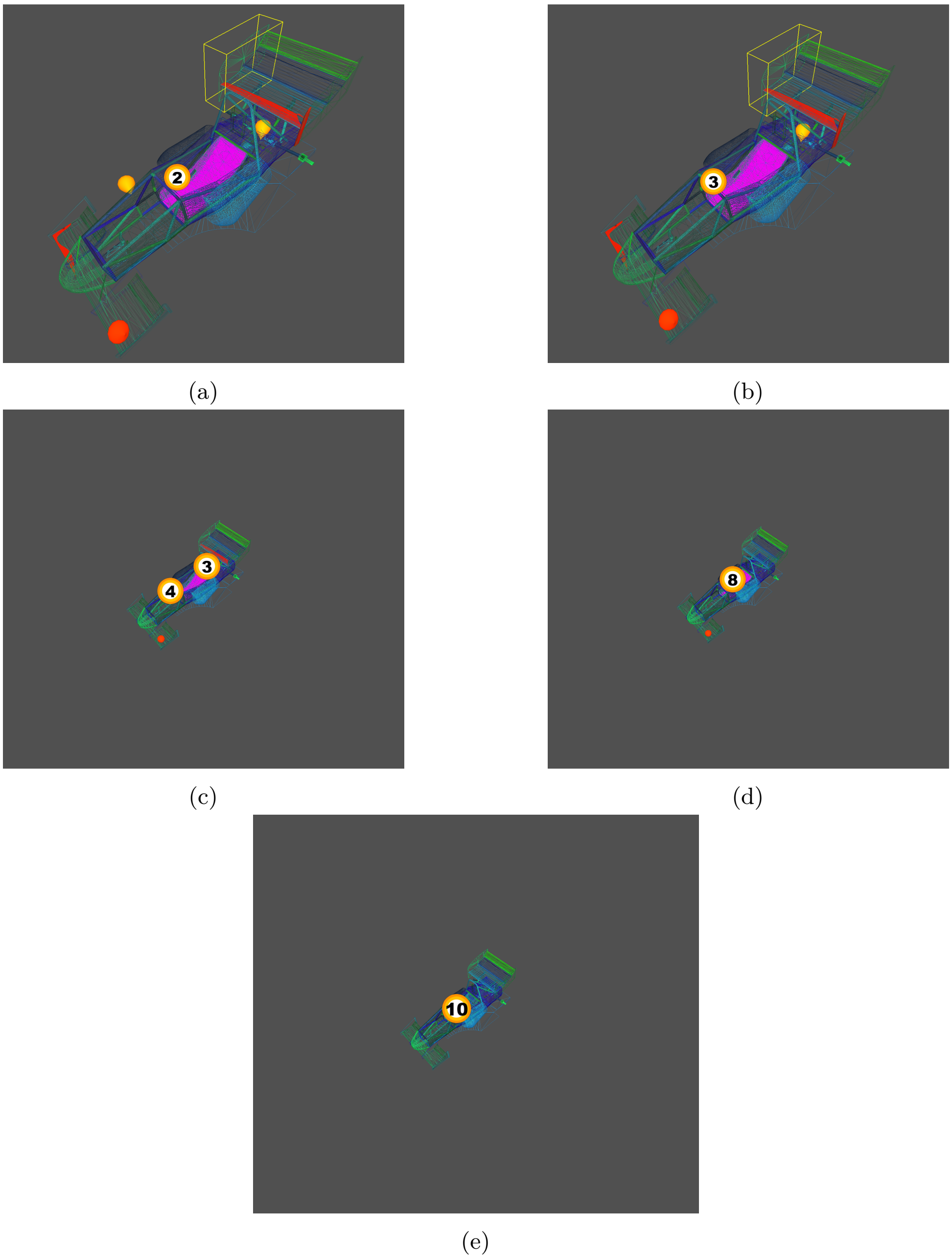
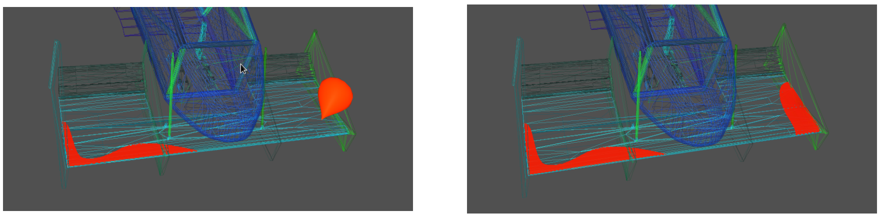


Figure 7.8: A Layer-Of-Detail strategy in action, showing how VIOs cluster and the user zooms out.



(a) FOI Clustered

(b) FOI Unclustered

Figure 7.9: FOI Clustered and Unclustered

- Component-Of-Interest, Section-Of-Interest, Feature-Of-Interest, and Point-Of-Interest are all needed.
- Participants would prefer the use of Components-Of-Interest for the display of information in model-based virtual environments.
- Region-Of-Interests, Sections-Of-Interests and Points-Of-Interest are also required in certain scenarios.
- There is some ambiguity between the use of directional and non-directional Points-Of-Interest.

In answer to the research question, information is related to the product in range of ways and each of these should be accounted for when visually displaying the information.

The last section of this chapter outlined a layer-of-detail strategy to manage the display of highly clustered VIOs and VIOs too small to be visible to the user. While not evaluated as part or the study presented here, the layer-of-detail strategy is however necessary and was included here for completeness.

Chapter 8

Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design

8.1 Introduction

The fourth and final Research Question: *How does model-based information navigation improve engineering information access and knowledge discovery?*, considers the affordance of model-based information navigation and, in particular, how navigating a document collection via visual representation of the product impacts the way engineers find information and discover knowledge. Answering this question involved a number of stages that accumulated in an A-B test involving a set of surrogate tasks and users to compare the model-based approach to a more traditional text-based search engine.

This chapter then describes the design of this A-B test that formed the final Descriptive Study 2 stage of the research methodology. It starts with the description of an analysis of the types of tasks performed by the Airbus In-Service team within their daily operations, these tasks where then mapped to Formula Student and literature such that the study is well-founded in both literature and the 'real world' of engineering organisations. The study itself is then described, followed by a description of the model-based and text-based systems used, the measurements taken, and the participants. The results from the study are described in the next chapter of the thesis: **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**.

8.2 Engineering Information Needs

The first stage in investigating the differences between a model-based and text-based approach to information navigation is determining the engineering information needs at the Airbus In-Service team use-cases. The In-Service team respond to repair requests form customers and these requests generate information needs. Understanding these needs and using them as a foundation for the study will ensure the findings are relevant to the Airbus use-cases that this thesis is founded on.

8.2.1 An Audit of the Engineering activities within In-Service Repair Request

The process of a repair request begins with an airline maintenance team either emailing or phone calling a regional contact centre. A form is then completed containing the relevant information that is forwarded on to an engineer/case manager. The process can take a number of directions that include providing a repair from a past repair to a full-blown design and testing of a new repair solution. The cycle ends with every document relating to the case being printed and then scanned to form one large PDF file that is electronically stored and made available via a search engine.

The Wing In-Service team provided 240 reports from the year 2013. Reports related to the single-isle A320 range of aircraft. For the purpose of the study, the engineering task(s) from each report were identified and extracted from the reports and are summarised in Table 9.11. Some reports contained a number of tasks, a typical example being where the maintenance team would suggest their own repair and ask the Airbus team to analyse damage, evaluate their proposed repair and if their repair was insufficient, provide repair instructions. As Table 9.11 shows, these three types of task were by far the most common.

ID	Activity	Frequency
1	analyse extend of damage	89
2	evaluate proposed repair	117
3	provide repair instructions	118
4	approval for use of new material in repair	2
5	order item	1
6	provide technical drawing	1
7	report manufacture issue	1
8	approve repair	38
9	instruct on inspection methods	1
10	provide aerodynamics analysis	2
11	provide information	9
12	search for similar incident	1

Table 8.1: A summary of the engineering activities contained within 240 repair queries received by the Airbus Wing In-Service function during 2013

8.2.2 Mapping Airbus Activity to Literature

There are a number of research areas associated with engineering information and the ways in which engineers use information that scientifically underpin the information needs and activities gathered from the In-Service team. Ward [153] discusses where engineers find information and lists sources of information such as memory, personnel files and department files and databases. The In-Service is in line with these findings with their main information source being the Daedalus search engine [8]. Similarly, Robinson [154] lists the cognitive skills of engineers as involving: locating information source, locating information within source, understanding information, problem solving and decision-making. Again, these are all skills that the In-Service personnel use as part of the response to a repair request with the main still being to locate sources and relevant past cases and the information within it.

For completion, Table 8.2 lists the engineers' knowledge and information needs published by Heisig [3] and maps these to the activities from the In-Service team shown in Table 9.11. This mapping shows how the In-Service knowledge and information needs cover a wide range of those published and from this it can be said that the In-Service team have knowledge and information needs that are consistent with the wider engineering field presented in literature.

To ensure the appropriateness of the study to evaluate the wider industrial use-cases, the Airbus activities identified in Table 9.11 are now translated to the world of Formula Student. Table 8.3 shows real-world Formula Student tasks that act as surrogates for the Airbus In-Service Tasks identified. These Formula Students tasks mimic the In-Service tasks such that they are a direct replacements that maintain the real world activities

Author	Information Need/Task	Airbus Activity ID											
		1	2	3	4	5	6	7	8	9	10	11	12
Heisig[3]	rationale												
	component/part				x	x		x	x	x			
	changes/modifications		x		x				x				
	drawing		x				x	x	x				
	design description	x	x		x		x		x				
	service									x			
	specification	x	x		x			x	x				
	requirements	x	x		x			x	x				
	difficulties/problems/issues	x	x					x		x			
	material				x								
	calculations/analysis	x	x								x		
	method			x						x		x	x
	manufacturing information		x	x				x					
	failures	x	x					x					
	test		x		x				x				
	constraints		x		x				x	x			
	decision				x				x				
	design documents	x	x		x		x	x	x				
	parameter												
	model(s)												
	product	x	x			x				x			
	costs				x	x			x				
	functions	x	x										
	design solutions		x		x				x				
	geometry	x	x				x						
	performance		x		x			x	x				
	software												
	feedback and suggestions												
	maintenance information	x								x			
	options and choices		x		x				x				
	reliability		x		x			x					
	input data	x	x		x						x		
	learnings	x	x										x
	design process		x										
	projects												x
	reports/records (non-design)					x							x
	supply chain							x					
	people												
	achievements												
	as-delivered	x	x				x						
	features		x										
	design for reuse												
	references												
	safety and risks		x		x				x				
	behaviour	x	x										
	best practice		x										
	end-user support			x									
	functional relationships	x											
	technologies		x										
	as-built info	x	x	x				x					
	assumptions		x										
	product end life												
	standards	x	x					x		x			
	timeline												
	plans												
	criteria		x		x				x				
	marketing												
	meeting minutes												
	organisation processes/structures			x	x				x				
	peers and competitors												
	sales												
	design reviews	x	x										
	resources		x										
	correspondence												
	patent		x										
	stakeholders			x									
	terminology/glossary/definition												

Table 8.2: Mapping engineering activities and information needs to those found in Heisig [3]

and information needs of the In-Service department. The purpose of doing this being to insure the proper use of the Formula Student as a surrogate for the Wing In-Service team.

ID	In-Service Activity	FS Activity	Summary of Equivalence
1	analyse extent of damage	analyse requirements (FS regulations)	both require the comparison of real world information to written standards
2	evaluate proposed repair	evaluate last year sub-system design	both require the evaluation of technical information
3	provide repair instructions	provide design instructions	directly equivalent
4	approval for use of new material in repair	approve for use of new material	directly equivalent
5	order item	order item	directly equivalent
6	provide technical drawing	provide technical drawing	directly equivalent
7	report manufacture issue	report manufacture issue	directly equivalent
8	approve repair	approve design	both require the analysis of technical information
9	instruct on inspection methods	instruct on assembly method	both require technical procedural information
10	provide aerodynamics analysis	provide aerodynamics analysis	directly equivalent
11	provide information	provide information	directly equivalent
12	search for similar incident	search for similar report	both require the searching through PDF documents and determining similarity base on technical information
13	identify most common repair case locations over set period of times	identify the area of the car containing the most information	both ask to analyse the document distribution across the car
14	identify most common repair case types over set periods of time	identify areas of the case worked on by an individual	both require a summation of content within a set of documents
15	identify new knowledge from a seemingly unconnected past case	identify new knowledge from a seemingly unconnected source	both require the searching through documents and determining similarity base on an information need

Table 8.3: List of the Airbus Wing In-Service activities mapped to the Formula Student surrogate activities

8.2.3 Formula Student Study Tasks

As described in **Chapter 4: Aim, Methodology and Research Questions**, the safety-critical nature of the worked performed by the Airbus In-Service team resulted in them being unable to participate in this final study of the thesis. As such, a number of Formula Student Racing teams, surrogate users, data sets and tasks were used. The Methodology chapter describes the appropriateness of using Formula Student in place of Airbus personnel through identifying the similarities in skills and engineering tasks performed by the two groups. Table 8.3 builds on this by mapping the Airbus Wing In-Service tasks to those performed by the Formula Student teams.

8.3 Study

This section now outlines the details of the study aimed at comparing the model-based approach to a more traditional text-based search engine to identify the affordancce of the model-based approach.

8.3.1 System Design

The study formed an A-B test, that is the two versions of the system, in this case a search engine, are directly compared to illicit the differences between them. In this case, the model-based user interface being compared to a control of a more traditional text-based interface. Performing an A-B test then required the construction of both a model-based user interface and a more traditional text-based search engine. The key challenge in doing this is ensuring that the implementation is developed such that the aspect being compared is the visual representation and the traditional method of search above all else. To achieve this certain design decisions were taken to 1) minimise the learning curve associated with the new model-based approach such that participants are not inhibited by the unfamiliar system and 2) maximising the similarity between the two systems such that the major difference is between the model-based interface and text-based interface to ensure that it is this difference that is being examined rather that any other aspect of the implementation.

As such, the decision was made to implement a 'minimal' version of the system that did not include the document classification based document indexing strategy outlined in **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing** and the VIO outlined in **Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation**. Instead, a traditional TF-IDF search engine was constructed following the procedure described in **Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments**. The model-based approach was then integrated with this search engine to create the heat map and document filtering functions outlined in **Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments**. Associating documents to the product structure involved creating a folder hierarchy that mimicked that of the product structure, a method discussed in the findings of **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**. Relevant documents for each level of the product structure are linked to the car components such that clicking on a component retrieves those documents.

All this ensured the dependent variable being measured was the user interface (text-based vs. model-based). Table 8.4 shows a list of independent variables and how they were accounted for.

Independent Variable	Description
Effect of learning	Random order of questions
Indexing technique	Both interfaces use the same index.
Presentation of results	Both interfaces use a side-ways scroll and an image of the front cover as a preview. Accessing report is done via clicking on the image.
Bias in the tasks	Tasks presented in a random order and random interface assignment to each task (a or b).
Effect of browser/operating system	Chrome Web browser and Window 7 or 10 used throughout. No difference between performance on Windows 7 or 10.

Table 8.4: The list of independent variables and their description.

Figure 8.1 shows the model-based user interface and Figure 8.2 shows the interface with a range of features labelled. From Figure 8.2 then:

Label 1: The panel added to manage interaction with the study questions, from left to right, the panel contains the Study ID, the question, a free text box for answer input, a 'Switch Representation' button and a 'Start/Next Question' button.

Label 2: The Product Structure List, a textual hierarchical list of the product structure. As the user clicks on the list, the visualisation changes to show just the selected parts of the car.

Label 3: The HTML5 Canvas that manages the display of the car model.

Label 4: The Navigation Controls panel containing the heat map search box, the onion peeling/MRI controls, and toggle buttons for the drag, hide heat map, exploded view, and reset mesh.

Figure 8.4 shows the interface to the text-based search engine. The study section of the interface remains consistent across the two representations. Figure 8.3 and 8.5 show the results pages from the model-based and text-based representations respectfully. These are designed to be as similar to each other as possible while maintaining consistency with the underlying representation. For example, both provide the front page of the report as a summary of the document and these are browsed via a left-to-right scroll. The model-based approach however, allows the user to still see the model-based environment in the background as a reminder that the search was being performed in a non-pure text environment.

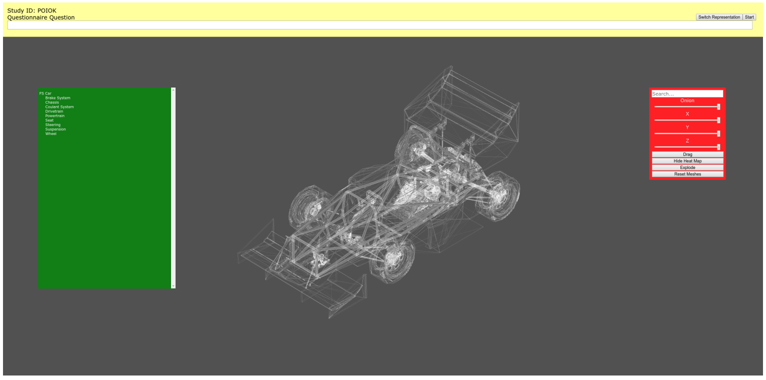


Figure 8.1: A screen shot of the Model-Based Information Navigation test system

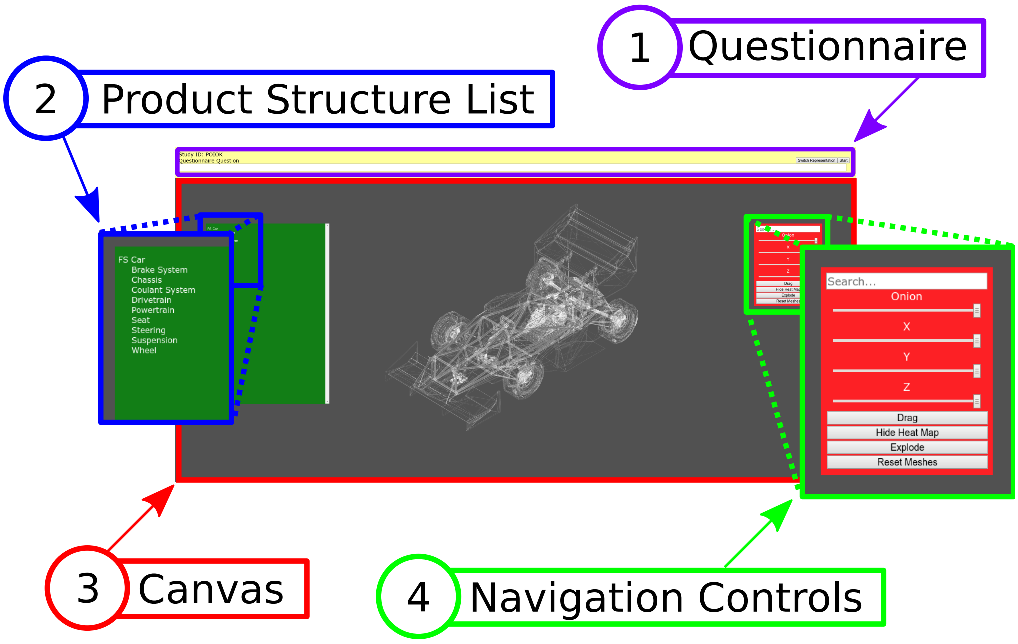


Figure 8.2: A screen shot of the Model-Based Information Navigation test system

8.3.2 Measures

Chapter 4: Aim, Methodology and Research Questions outlined the range of measures typically used in evaluating the system outlined in this thesis. Given the real world nature of the use-cases outlined in this thesis, it is appropriate that the measures

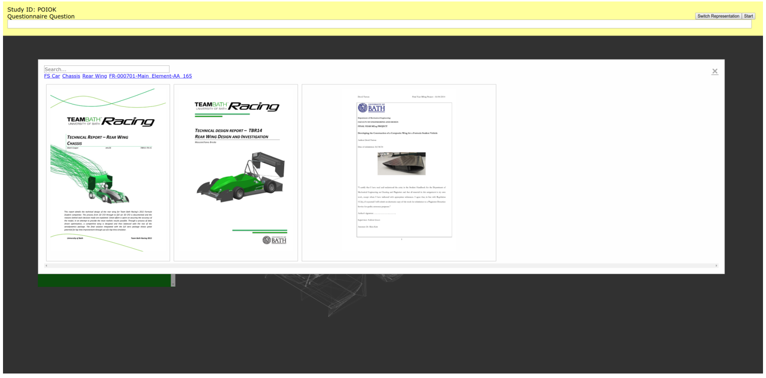


Figure 8.3: A screen shot of the results page of the Model-Based Information Navigation test system with additional navigation controls and questionnaire panel

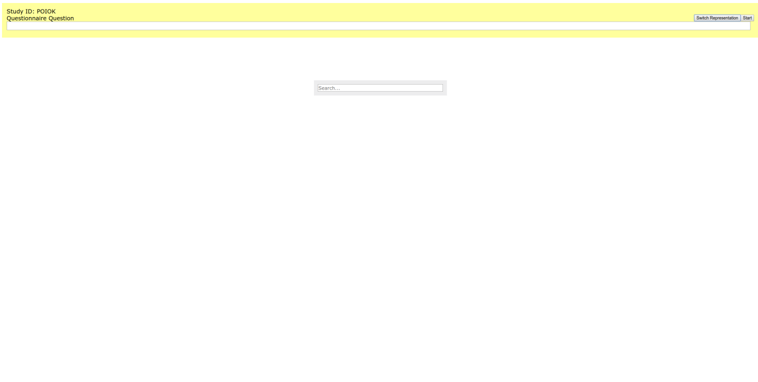


Figure 8.4: A screen shot of the text-based search engine test system

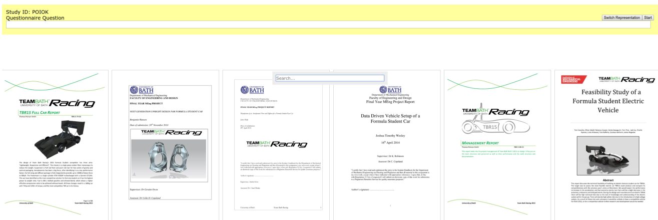


Figure 8.5: A screen shot of the results page of the text-based search engine test system

are also real-world in nature. This study then primarily focused on the user's ability to find an answer to an information need, the time taken to find an answer, and the user's preference for the model-based system against the text-based system.

To evaluate information seeking success and speed, the system captured the user interactions with the system (mouse movements, mouse clicks, key presses, queries entered and reports opened for examples). Along with these events, the time of each event

```
Study ID: NFTQL
Tue Mar 20 2018 15:38:34 GMT 0000 (GMT Standard Time)
0: Name a past supplier of sprocket carriers.
click,x:1877,y:54, 1017493.1999999972,[object HTMLInputEle-
ment],questionnaireNextButton
click,x:1877,y:54, 1017493.1999999972,[object HTMLInputEle-
ment],questionnaireNextButton
mousemove,x:1873,y:54, 1018517.2999999996,[object HTMLInputEle-
ment],questionnaireNextButton
mousemove,x:1869,y:56, 1018565.300000002,[object HTMLInputEle-
ment],questionnaireNextButton
mouseout,x:1841,y:68, 1018613.1999999997,[object HTMLInputEle-
ment],questionnaireNextButton
mouseover,x:1841,y:68, 1018613.1999999997,[object HTMLInputEle-
ment],questionnaireTextInput
mousemove,x:1841,y:68, 1018613.1999999997,[object HTMLInputEle-
ment],questionnaireTextInput
mousemove,x:1769,y:90, 1018653.2999999982,[object HTMLInputEle-
ment],questionnaireTextInput
mouseout,x:1637,y:116, 1018701.3000000006,[object HTMLInputEle-
ment],questionnaireTextInput
mouseover,x:1637,y:116, 1018701.3000000006,[object HTMLDivElement],textResultsDiv
mousemove,x:1637,y:116, 1018701.3000000006,[object HTMLDivElement],textResultsDiv
```

Figure 8.6: An example of a logging file generated from interactions with the test system

was also recorded, allowing for the recording of both total time and the time spent performing any particular activity. See Figure 8.6 for an example of the beginning of a log file. The figure shows the first line stores the Study ID, followed by the date and time and then the question being answered.

To capture user preference, a paper-based questionnaire based on the IBM Computer System Usability Questionnaire (CSUQ) [79] was created. The questionnaire consists of a number of Likert Scale type questions, asking the user to agree/disagree with a statement using a 1-7 scale with an N/A option and a final request to comment on the answer given.

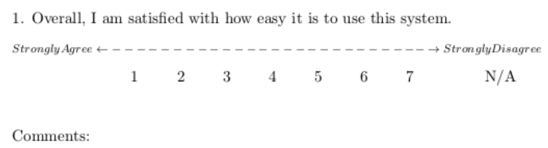


Figure 8.7: An example of a question from the IBM Computer System Usability Questionnaire

Unadulterated, the questionnaire covers all aspects of user preference for a computer system, including access to help files and system settings. These were not provided to participants during the study and so the relevant section of the questionnaire was not used. The questionnaire was also adapted to include participant demographics and five additional questions aimed at capturing the participant's preference for one system over the other. See Appendix A.2 for the questionnaire.

The qualitative and quantitative measures were linked via the 'Study ID' shown at the top of the screen (see Figure 8.2) which is captured in the log file and in the first page of the questionnaire.

8.3.3 Participants and Locations

Formula Student teams from the University of Bristol, University of Bath, University of West England and Imperial College London participated in the study over four separate days situated at general computer teaching rooms at their own institution. The computers used either Microsoft Windows 7 or Windows 10 and students were instructed to use the Google Chrome Web-browser to access the study website. The size of the groups ranged from Imperial College London with four participants to the University of Bristol with 19 participants. Participants also performed a range of roles within their team, ranging from first year undergraduate CAD engineers to final year masters level team leaders. A range of disciplines were also covered, ranging from aerospace engineers to electrical, mechanical and automotive. Table 8.5 describes the participants in greater detail.

Study Participants	
UWE	14
Bath	6
Bristol	19
Imperial	4

Table 8.5: A summary of the engineering tasks contained within 240 repair queries received by the Airbus Wing In-Service function during 2013

The UWE participants were used to test and provide feedback on the system. As stated in **Chapter 6: Navigation in Model-Based Three-Dimensional Engineering Virtual Environments**, two changes were made following feedback from this group: a feature that collapsed the product structure list in the green panel on the left of the screen, and the ability to zoom in and out. It was decided that the UWE participants' results should be included in the total set of results presented in **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**. The justification for this was that the changes to the interface would have a minimal effect on performance, and that this effect would be 1) small compared to the effects of the dependent variable (i.e. the interface (model-based and text-based)) and 2) largely mitigated by averaging results across all universities. This decision acknowledged that the UWE study was undertaken on a slightly different system to that of the wider study, and that this should be considered when reviewing the final results.

8.3.4 Tasks

The final aspect of the study design involved generating the tasks that participants will be engaged in. Table 8.6 shows a list of nine pairs (a and b) of questions that were designed to invoke one or more of the Formula Student information tasks described in Table 8.3. For each of the nine pairs, the two questions were randomly assigned to one of the interfaces to eliminate any unintended question bias. The order of the questions were also randomised for each participant to eliminate the effect of learning.

The study was first performed at the University of West England and these results were initially treated as a test of both the system and the study. The outcome of this study resulted in a number of small changes being made to the interface and questions 4.c and 9.c being replaced by questions 4.b and 9.b respectfully. The user interface changes added zoom functionality to the model-based interface and a 'collapsible' feature to the product structure list. Questions 4.c and 9.c were replaced with 4.b and 9.b because observations showed that, in comparison with the other questions, participants were spending a relatively large amount of time searching through the reports for the information.

8.3.5 Study Structure

During each study, approximately 15-20 minutes was spent introducing the participants to the project and demonstrating both the model-based user interface and the text-based interface. Participants were then given 15 minutes to familiarise themselves with the system such that they were confident in their ability to use both interfaces to find information. The 'Switch Representation' button (see Figure 8.2 allow them to move between the two interfaces. After this, the students were instructed to move on to the study itself.

An hour and a half was allocated for the study and participants were asked to complete a paper-based questionnaire on completion. Questionnaires and system logs were linked via a unique Study ID that was displayed on the screen at all times and captured by participants in the first part of the questionnaire. Those participants who had not completed within the allocated time were asked to stop and complete the questionnaire.

In the event that the study was interrupted, participants were told to make a note of the old Study ID and to restart. This essentially adds a level of variability in the number of questions answered and introduces cases where the user answers the same question more than once. The justification for allowing this is that it represents the real-world scenarios of finding and re-finding information in busy work places environments i.e. the Airbus Wing In-Service team. Students were told to make their own judgement as to whether they could answer a question, and respond with either the answer or a statement saying it was not possible to find the information.

8.4 Discussion and Conclusion

This chapter described the design of an A-B test to compare a model-based and text-based approach to information access and knowledge discovery with the aim of determining the whether a model-based approach to information navigation improves information access and knowledge discovery. Described was the analysis of the types of tasks performed by the Airbus In-Service team, these tasks where then mapped to Formula Student tasks. Formula Student being the surrogate use-case for the study. Following this, a description of the model-based and text-based systems used, the measurements taken, and the participants was presented.

The next chapter of the thesis **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**, presents the results from the study followed by a more complete discussion and conclusion.

Number	Question	Task	ID
1.a	Can you find the document '2018 Supplementary Rules' by the Institution of Mechanical Engineers?	analyse requirements (FS regulations)	1
1.b	Can you find the textbook called 'Advanced Brake Technology' by Bert Breuer and Uwe Dausend?	analyse requirements (FS regulations)	1
2.a	Do you think that past teams have completed sufficient research into tyres such that suitable tyres could be specified for the current car without additional research?	evaluate last year sub-system design; approve design	2, 8
2.b	Do you think that past teams have completed enough research into impact attenuators that a past year's design could be re-used this year?	evaluate last year sub-system design; approve design	2, 8
3.a	Name the area(s) of the car that have received the most computational fluid dynamics analysis.	identify the area of the car containing the most information	13
3.b	Name the area(s) of the car that have received the most finite element analysis.	identify the area of the car containing the most information	13
4.a	Name a past supplier of sprocket carriers.	order item	5
4.b	Name a past supplier of electric motors.	order item	5
4.c	Name a past supplier of brake pads.	order item	5
5.a	Has anyone explored the use of additive manufacturing for the tripod housing?	provide information; approve for use of new material	11, 4
5.b	Has anyone explored the use of additive manufacturing for the upright manufacture?	provide information; approve for use of new material	11, 4
6.a	Find a front wing general assembly drawing. Did the author include their name and if so, who created the drawing?	provide technical drawing; instruct on assembly method; provide design instruction	6, 9, 3
6.b	Find the main hoop technical drawing. Did the author include their name and if so, who created the drawing?	provide technical drawing; instruct on assembly method; provide design instruction	6, 9, 3
7.a	Find a report on active aerodynamics. What is the file called?	provide aerodynamics analysis; search for similar report	10, 12
7.b	Find a report on the use of analysis data to improve performance. What is the file called?	provide aerodynamics analysis; search for similar report	10, 12
8.a	What area(s) of the car did [redacted] work on?	identify areas of the case worked on by an individual	14
8.b	What area(s) of the car did [redacted] work on?	identify areas of the case worked on by an individual	14
9.a	Who would you contact for advice about carbon fibre wheel rims?	identify new knowledge from a seemingly unconnected source; report manufacture issue	15, 7
9.b	Who would you contact for advice about front inboard suspension?	identify new knowledge from a seemingly unconnected source; report manufacture issue	15, 7
9.c	Who would you contact for advice about rear outboard suspension?	identify new knowledge from a seemingly unconnected source; report manufacture issue	15, 7

Table 8.6: A summary of the engineering tasks contained within 240 repair queries received by the Airbus Wing In-Service function during 2013. Names of individuals have been redacted for anonymity.

Chapter 9

Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results

9.1 Introduction

In the previous chapter (**Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**), a study was outlined that aimed to answer the fourth and final Research Question: *How does model-based information navigation improve engineering information access and knowledge discovery?* This chapter presents the results from the study. These are presented on a question-by-question basis before some general results, from across all questions are discussed. The results are presented per question and include figures showing the total time to complete the question, a breakdown of statistics relating to the total time, a graph showing the search, navigation, and browsing tasks performed within each question, a graph showing the participant success rate in being able to find an answer to each question, a Venn diagram showing the total number of reports opened in response to the questions, and a selection of relevant questionnaire comments.

9.2 Measure Calculations

Outlined in **Chapter 4: Aim, Methodology and Research Questions** and **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** were a range of measures aimed at evaluating Information Retrieval from a human-computer interface perspective. The measures presented here are:

Measure	Description
Total time for completion	The task level measure of total time to complete a question.
Comparison of activity	A breakdown on the percentage time spent on navigation, searching and browsing activities during each task. (Will only be direct comparable when there is no significant difference between total time).
Success rate	Whether the participant could find an answer or not. Participants were asked to state if they could not find an answer.
Number of Reports Opened	A measure of effort required to complete the task and the effectiveness of the search engines/ranking. I.e., the more reports opened, the more effort required and the less effective the search engine/ranking.
Selection of Relevant Comments	A reflection on the participants option on the aspect of the interfaces/study that are relevant to the task/results presented.

Table 9.1: The list of measures presented from the A-B study.

The total time to complete questions were determined using the system interaction logs. The difference between the first mouse-move event and the final mouse-over the 'Next Question' button were taken. The presented results have outliers beyond two standard deviations removed. Statistics on the remaining values are then performed and presented. A Shapiro-Wilks test was used to determine skewness of both the model-based and text-based results and these results then determined the significance test method - a Paired T-Test where both set of results are normally distributed and a Wilcoxon Test when they were not. A significance (p-value) of 0.05 is used throughout.

The search, navigation and browsing activities were calculated by classifying the range of activities generated within the logs. Time spent performing activities relating to the task or those that could not be classified (participant move the mouse off-screen) were ignored. Traditional search breaks down into time spent formulating queries and time spent browsing results. The purpose of exploring these results is to determine the effect that adding a stage (navigating the three-dimensional model) has on the search.

The success rates were determined by whether the participant noted that they could not find an answer (as instructed). No evaluation was made on the validity of the answers themselves, the rationale being that the participants are the domain experts and as such could judge themselves if an appropriate answer could be found. So, if an answer was provided, it was deemed a successful search.

The number of reports access was also explored in an attempt to judge whether the model-based interface reduces the search space, i.e. reduces the number of reports that participants had to view. These are shown in a Venn diagram with the intersect showing those reports opened during searches with both interfaces.

In addition to the results for each question, a selection of relevant comments were extracted from the questionnaire in an attempt to aid explanation of the results.

9.3 Results

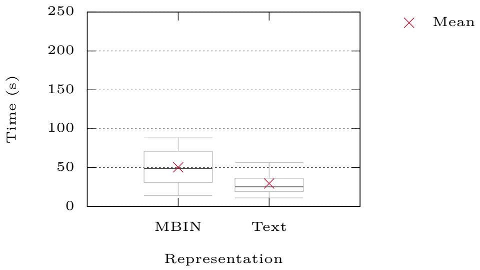
This section now presents the results, it starts by presenting each task separately before presenting a summary of all tasks, demographics and finally a range of participant comments.

9.3.1 Task 1: Analyse Requirements

See Table 9.2 for the questions relating to this task. Figure 9.1 shows a range of results in response to the first pair of questions. Figure 9.1a shows the distribution of times taken to answer the question. The statistics presented in Table 9.1b show the results for both samples are normally distributed and as such, a Paired T-Test is used to determine the significance between the total times for the two interfaces. There is a significant decrease in time of around 20 seconds for the text-based interface compared to the model-based interface. Exploring this further, Figure 9.1c shows the differences in percentage time spent on search, navigation and browsing activities, this graph shows that approximately the same amount of time is spent navigating and browsing when using the model-based interface. Around 80% of text-based activities are in searching or formulating queries. The final Figure 9.1d, shows how, when using the model-based interface, participants failed to find the answer half of the time, compared to around 20% of the time with the text-based system. The Venn diagram depicted in Figure 9.1e shows the total number of different reports opened by all participants when browsing for the answers. The diagram shows how participants using the model-based interface opened over twice the number of reports (11 vs 5) as the text-based interface and only two of those reports were opened during use of both interfaces. The comments from the questionnaire (Table 9.1f) highlight how participants could not find the brake pads in the model-based interface, and although participants stated they prefer the model-based interface for component specific information, they also found it difficult to locate the '*2018 Supplementary Rules*' document. A document that is not obviously tied to any one particular component.

	Questions
a	Can you find the document '2018 Supplementary Rules' by the Institution of Mechanical Engineers?
b	Can you find the textbook called 'Advanced Brake Technology' by Bert Breuer and Uwe Dausend?

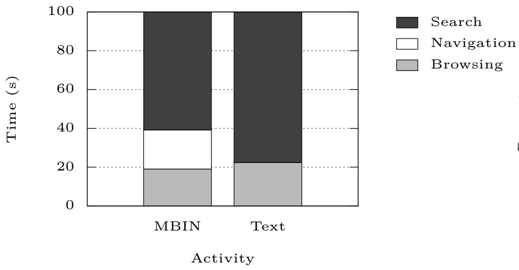
Table 9.2: The questions from Task 1



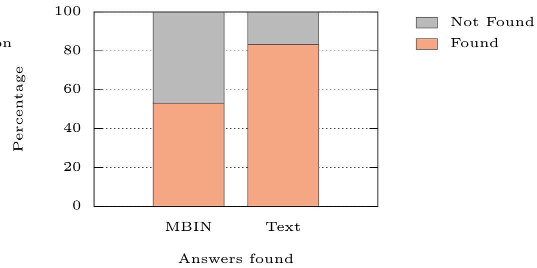
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	22.00	22.00
Mean	50.29	29.57
Standard Deviation	23.23	13.07
Skew	-0.07	0.57
Skew Standard Error	0.49	0.49
Kurtosis	-1.43	-0.81
Kurtosis Standard Error	0.95	0.95
Shapiro-Wilk	0.92	0.93
- p-value	0.07	0.13
Significance		
T-Test		3.40
- p-value		0.00

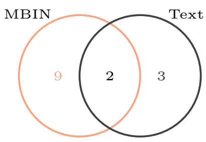
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*Brake pads appeared to be missing*'
- '*Could be very useful for a lot of things, but it's hard to find some general documents i.e. the book on brakes - I found that after answering another question.*'
- '*Usually, some searches (e.g. rules) are easier in text based.*'
- '*System works well to find files on specific components but not for general car studies or non-component areas.*'
- *Definitely more efficient for finding information on particular components...*'

(f) Selection of relevant questionnaire comments

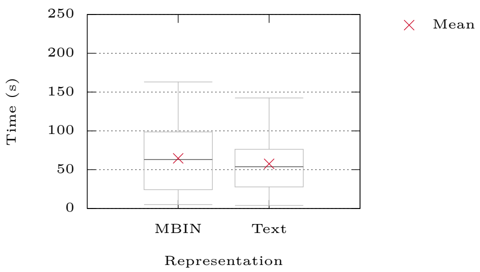
Figure 9.1: Results from Task 1

9.3.2 Task 2: Evaluate Last Year Sub-System Design; Approve Design

See Table 9.3 for the questions relating to this task. Figure 9.2 shows a range of results in response to the second task with Figure 9.2a and Table 9.2b showing no significant difference between the time taken to complete the questions. The statistics presented in Table **??** show the results for both samples are normally distributed and as such, a Paired T-Test is used to determine that there is no significance between the total times for the two interfaces. Given the similarity between the time to complete the question, it is possible to see, from Figure 9.2c, how the time spent browsing results is similar between the two interfaces and that time spent navigating the model in the model-based interface reduces the amount of time spent formulating search queries compared to the text-based interface. Figure 9.2d shows how again, participants were less likely to find results using the the model-based interface, however, for this question the success rate for the model-based interface is higher at approximately 80%. Figure 9.2e shows how participants needed to open 21 and 27 reports to answer the questions using the model-based and text-based respectfully, with four of those in common between. So in this case, the participants opened fewer reports using the model-based interface compared to the text-based interface.

	Questions
a	Do you think that past teams have completed sufficient research into tyres such that suitable tyres could be specified for the current car without additional research?
b	Do you think that past teams have completed enough research into impact attenuators that a past yearś design could be re-used this year?

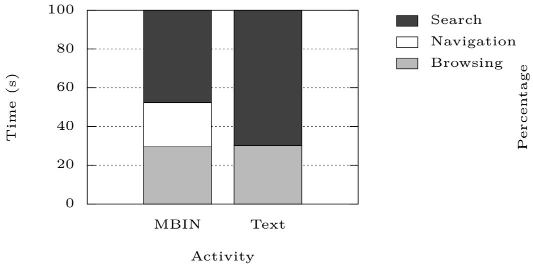
Table 9.3: The questions from Task 2



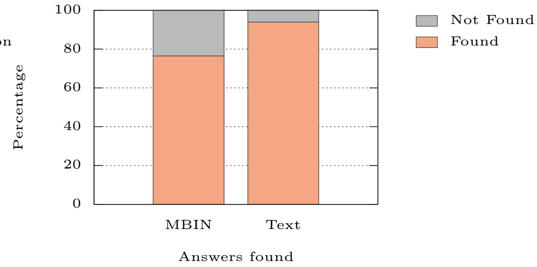
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	26.00	26.00
Mean	64.59	57.54
Standard Deviation	45.65	37.56
Skew	0.58	0.69
Skew Standard Error	0.46	0.46
Kurtosis	-0.71	-0.48
Kurtosis Standard Error	0.89	0.89
Shapiro-Wilk	0.93	0.93
- p-value	0.07	0.07
Significance		
T-Test		0.70
- p-value		0.49

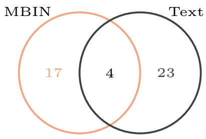
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments

(f) Selection of relevant questionnaire comments

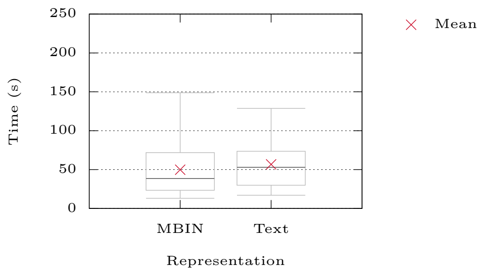
Figure 9.2: Results from Task 2

9.3.3 Task 3: Identify the Area of the Car Containing the Most Information

See Table 9.4 for the questions relating to this task. Figures 9.3 show the results for Task 3. Figure 9.3a compares the total times with the model-based interface approximately 7 seconds quicker, the statistics presented in Table 9.3b however show this time difference is not significant. The Shapiro-Wilks p-values is below the below 0.05 required to indicate a normally distributes set of results and as such, a Wilcoxon test was performed to determine the non-significance. Figure 9.3c shows how time spent browsing is fairly similar between the interfaces and that time spent navigating reduces the time spent formulating queries. Figure A.14 shows how participants were less likely to find an answer using the model-based interface compared to the text-based interface. Figure 9.3e shows the number of reports opened during the use of each interface and that between the interfaces and questions, only one report was shared. While the number of reports is similar, it is worth noting that it should have been possible to answer this question using the heat map and so not needing to open any reports. Participants opened fewer reports using the model-based interface compared to the text-based interface. The comments from the questionnaire discuss the use of the heat map - the technique designed to aid the answering of the knowledge discovery tasks. They show the heat map was well-received although one participant identified an error in the implementation where documents in the heat map responded to all terms in the document and not just the main topic discussed. On informally discussing the heat map with participants from Imperial College London post-study, two participants mentioned how one has to trust the heat map completely to quickly answer these questions, this may have result in fewer people relying on the heat map to complete this task.

	Questions
a	Name the area(s) of the car that have received the most computational fluid dynamics analysis.
b	Name the area(s) of the car that have received the most finite element analysis.

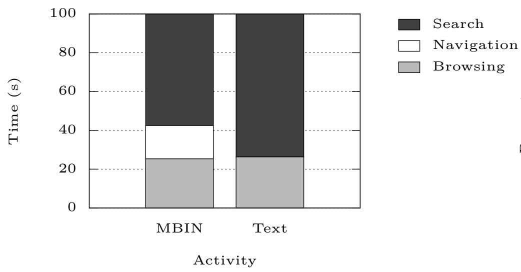
Table 9.4: The questions from Task 3



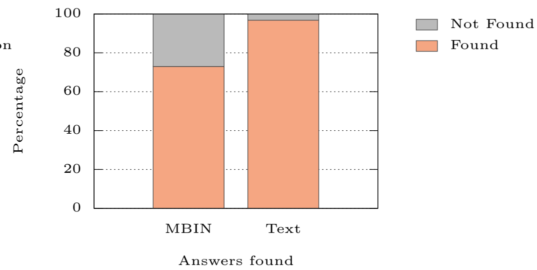
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	27.00	27.00
Mean	49.75	56.76
Standard Deviation	34.00	32.01
Skew	1.15	0.78
Skew Standard Error	0.45	0.45
Kurtosis	0.65	-0.32
Kurtosis Standard Error	0.87	0.87
Shapiro-Wilk	0.86	0.91
- p-value	0.00	0.02
Significance		
Wilcoxon		147.0
- p-value		0.31

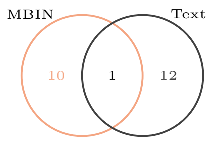
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*...the 3D model heat map displayed information well.*'
- '*The visual aide and heat maps are useful when searching for items and understanding the overall layout of the car.*'
- '*Love the drag function and heat map.* '
- '*...heat map can be misleading if documents talk about something but are not about that thing.*'

(f) Selection of relevant questionnaire comments

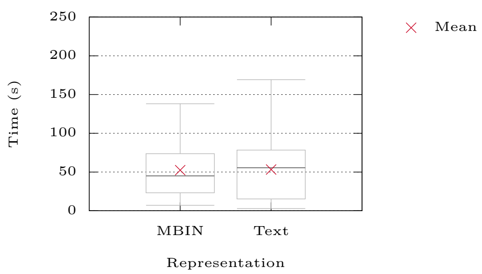
Figure 9.3: Results from Task 3

9.3.4 Task 4: Order Item

See Table 9.5 for the questions relating to this task. Figures 9.4 show the results for Task 4. Figure 9.4a and Table 9.4b shows no significant differences in the time to complete the questions. The statistics presented in Table 9.4b show the results for both samples are not normally distributed and as such, a Wilcoxon test is used to determine that there is no significance between the total times for the two interfaces. The similarity between the completion time means a direct comparison of the search, navigation and browsing activities can be made. Figure 9.4c shows how the time spent formulating search queries is similar, with time spent navigating reducing the time spent browsing results for the model-based interface. Figure 9.4d shows how the success rates for the two interfaces are approximately the same. The Venn diagram in Figure 9.4e shows how participants opened a smaller range of documents using the model-based interface compared to the text-based interface, with seven of those in common. The questionnaire comments show one user could not find the brake pads in the model-based interface while another found the model-based interface aided the finding of information in an unfamiliar area of the car.

Questions	
a	Name a past supplier of sprocket carriers.
b	Name a past supplier of electric motors.
c	Name a past supplier of brake pads. (UWE only)

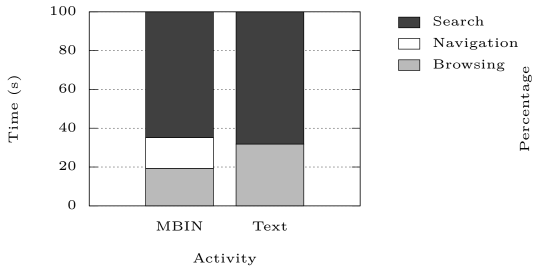
Table 9.5: The questions from Task 4



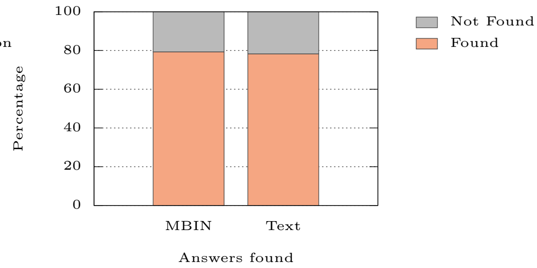
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	22.00	22.00
Mean	52.32	53.33
Standard Deviation	36.86	40.43
Skew	1.03	0.88
Skew Standard Error	0.49	0.49
Kurtosis	0.02	0.73
Kurtosis Standard Error	0.95	0.95
Shapiro-Wilk	0.87	0.91
- p-value	0.01	0.04
Significance		
Wilcoxon		126.0
- p-value		0.99

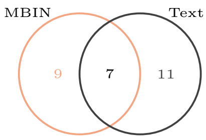
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*Brake pads appeared to be missing*'
- '*...to find some components it was easier to search because of my limited knowledge of the car parts.*'

(f) Selection of relevant questionnaire comments

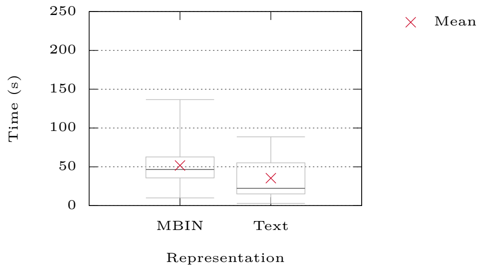
Figure 9.4: Results from Task 4

9.3.5 Task 5: Provide Information; Approve for Use of New Material

See Table 9.6 for the questions relating to this task. Figures 9.5 show the results form Task 5. Figure 9.5a and Table 9.5b show there is no significance in the quicker total completion time of the text-based interface compared to the model-based interface. The statistics presented in Table 9.5b show the model-based interface results are normally distributed however the text-based results are not. This results in a Wilcoxon text being performed to determine the non-significance. The differences between the mean time to complete does not allow for direct comparison between the percentage activities however, for the model-based interface the time spent navigating is once again similar to the time spent browsing results. Figure 9.5d shows how participants were less likely to find an answer using the model-based interface compared to the text-based interface. The Venn diagram (Figure 9.5e) shows how twice the number of reports were opened using the model-based interface compared to the text-based interface. One relevant questionnaire comment states how the participant struggled to find information related to an unfamiliar area of the car using the model-based interface and as such, preferred the text-based interface.

	Questions
a	Has anyone explored the use of additive manufacturing for the tripod housing?
b	Has anyone explored the use of additive manufacturing for the upright manufacture?

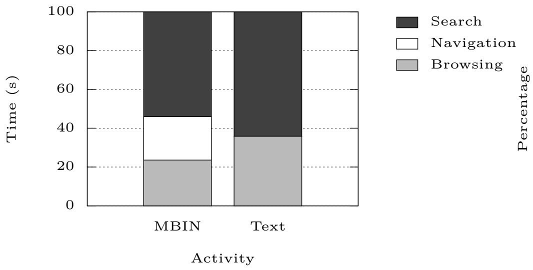
Table 9.6: The questions from Task 5



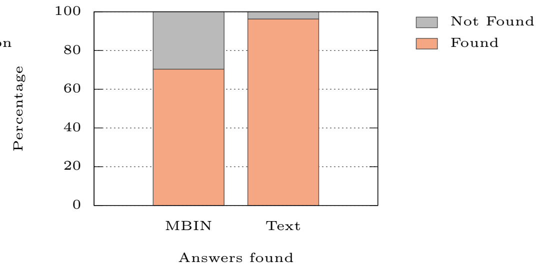
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	21.00	21.00
Mean	51.68	35.35
Standard Deviation	27.77	24.07
Skew	1.24	0.74
Skew Standard Error	0.50	0.50
Kurtosis	1.90	-0.72
Kurtosis Standard Error	0.97	0.97
Shapiro-Wilk	0.91	0.89
- p-value	0.06	0.02
Significance		
Wilcoxon		67.0
- p-value		0.09

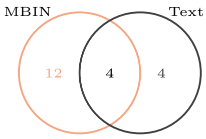
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*...to find some components it was easier to search because of my limited knowledge of the car parts.*'

(f) Selection of relevant questionnaire comments

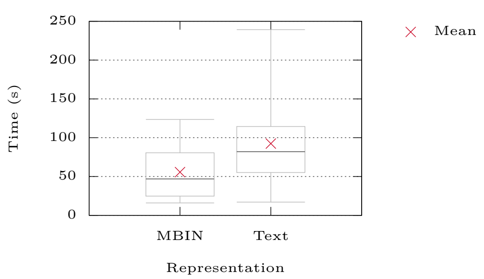
Figure 9.5: Results from Task 5

9.3.6 Task 6: Provide Technical Drawing; Instruct on Assembly Method; Provide Design Instruction

See Table 9.7 for the questions relating to this task. Figures 9.6 show the results for Tasks 6. Figure 9.6a and Table 9.6b show an approximate 40 second difference between the two interfaces, a significant improvement in the search time when using the model-based interface compared to the text-based interface. The statistics presented in Table 9.6b show the model-based results are normally distributed and the text-based are not and as such, a Wilcoxon test is used to determine that the significance of the 40 seconds difference. Figure 9.6c shows a similar distribution of time spent in search, navigation and browsing activities. Figure 9.6d shows that participants found it difficult to find the information using both interfaces, with only approximately 20% of participants finding an answer with the text-based interface and approximately 60% using the model-based interface. The Venn diagram (Figure 9.6e) re-enforces this difficulty, finding that 32 documents were opened using the text-based interface, 14 more than the model-based interface. The relevant questionnaire comments re-enforce this with many participants mentioning the difficulty in finding drawings and requesting/suggesting separate drawing search tools. It is worth noting here that the indexing interface was unable to extract text from the drawing and as such it was not possible to find the drawing using the text-based interface.

	Questions
a	Find a front wing general assembly drawing. Did the author include their name and if so, who created the drawing?
b	Find the main hoop technical drawing. Did the author include their name and if so, who created the drawing?

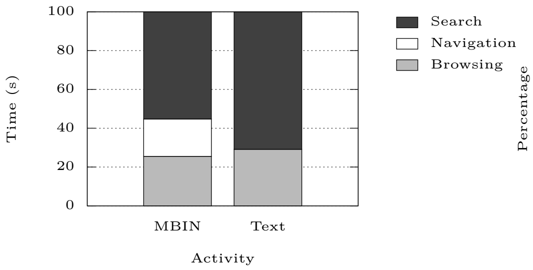
Table 9.7: The questions from Task 6



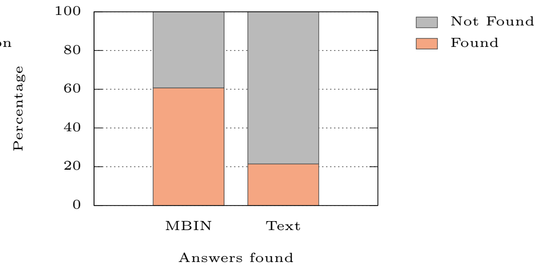
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	24.00	24.00
Mean	55.71	92.49
Standard Deviation	33.01	55.85
Skew	0.57	0.86
Skew Standard Error	0.47	0.47
Kurtosis	-0.86	0.08
Kurtosis Standard Error	0.92	0.92
Shapiro-Wilk	0.90	0.92
- p-value	0.03	0.06
Significance		
Wilcoxon		35.0
- p-value		0.00

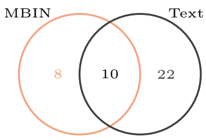
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*I found it hard to find assembly drawings.*'
- '*I had problems finding the drawings and usually everything that I search for was last and the less relevant ones were much earlier.*'
- '*It could have a direct link to a drawing when I click on a part.*'
- '*It would be great if technical drawings were directly linked to the 3D model and it would be possible to choose between drawing or documents.*'
- '*Searching for drawings is rather difficult.*'
- '*Also, a separate drawing tool would be useful.*'
- '*I would search drawing and no drawings would come up*'

(f) Selection of relevant questionnaire comments

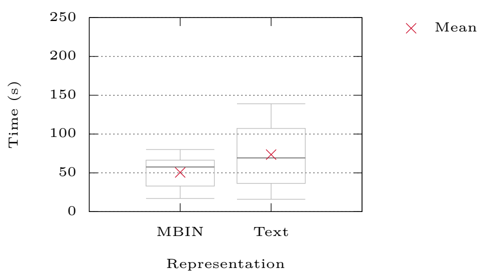
Figure 9.6: Results from Task 6

9.3.7 Task 7: Provide Aerodynamics Analysis; Search for Similar Report

See Table 9.8 for the questions relating to this task. Figures 9.7 show the results for Task 7. Figure 9.7a and Table 9.7b shows an approximate 20 second reduction in question completion time in for the model-based interface over the text-based interface. The statistics presented in Table 9.7b show both sets of results are normally distributed and as such a Paired T-Test is used to show the 20 seconds is significant. This total time difference does not allow for a direct comparison of the activities shown in Figure 9.7c. The question answering success rates shown in Figure 9.7d show participants were more likely to find answers using the text-based interface with an approximate 10% improvement over the model-based interface. The Venn diagram (Figure 9.7e) show the participants opened fewer reports using the model-based interface compared to the text-based interface. In spite of the improved overall completion time when using the model-based interface, the relevant questionnaire comments show some users found it more difficult to find information with the model-based interface when information/a document has no obvious link to a specific component.

	Questions
a	Find a report on active aerodynamics. What is the file called?
b	Find a report on the use of analysis data to improve performance. What is the file called?

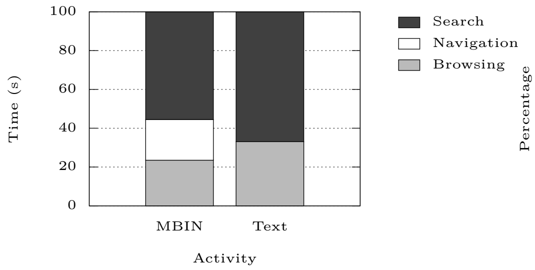
Table 9.8: The questions from Task 7



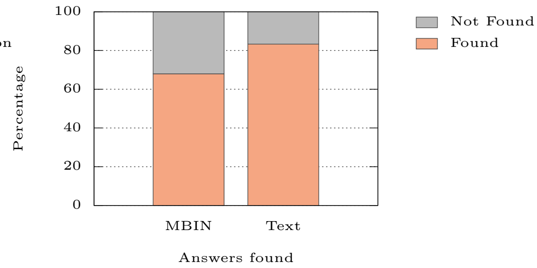
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	22.00	22.00
Mean	50.60	73.48
Standard Deviation	20.06	38.12
Skew	-0.28	-0.04
Skew Standard Error	0.49	0.49
Kurtosis	-1.22	-1.35
Kurtosis Standard Error	0.95	0.95
Shapiro-Wilk	0.93	0.93
- p-value	0.11	0.12
Significance		
T-Test		-2.46
- p-value		0.02

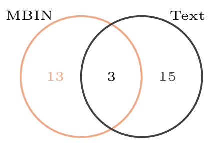
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*Simulation studies also don't really fit anywhere in the model, which makes them hard to find.*'
- '*however the 3D model is often confusing when you have to search something that doesn't have a specific part*'

(f) Selection of relevant questionnaire comments

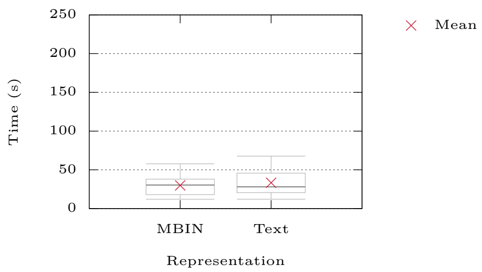
Figure 9.7: Results from Task 7

9.3.8 Task 8: Identify Areas of the Case Worked on by an Individual

See Table 9.9 for the questions relating to this task. Figures 9.8 show the results in response to Task 8. Figure 9.8a and Table 9.8b show no significant difference in the mean times to complete the questions using each interface. The statistics presented in Table 9.8b show both sets of results are normally distributed and as such a Paired T-Test is used to show there is no significance difference in the total times. The similarity in mean time to complete again allows for the direct comparison between search, navigation and browsing activities using Figure 9.8c. This graph shows a similar time spent browsing results with the time spent navigating in the model-based interface accounting for the time spent formulating search queries in the text-based interface. Figure 9.8d shows the success rates in answering the questions, with participants being able to answer the questions approximately 90% of the time with the model-based interface and 100% with the text-based interface. Figure 9.8e, the Venn diagram, shows the participants opened over twice the number of reports using the text-based interface compared to the model-based interface and that no common reports were opened. It is again worth noting that this question could be answered using the heat map in the model-based interface without opening any documents. The relevant questionnaire comments (Table 9.8f) show two participants felt the heat map allowed them to quickly view relevant areas of the car.

	Questions
a	What area(s) of the car did [redacted] work on?
b	What area(s) of the car did [redacted] work on??

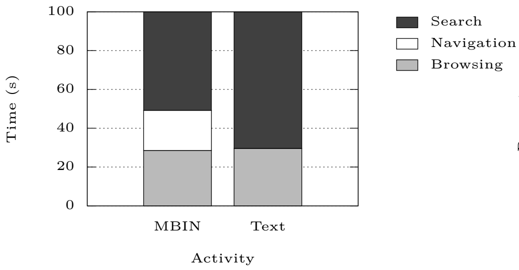
Table 9.9: The questions from Task 8 with individuals names redacted for anonymity



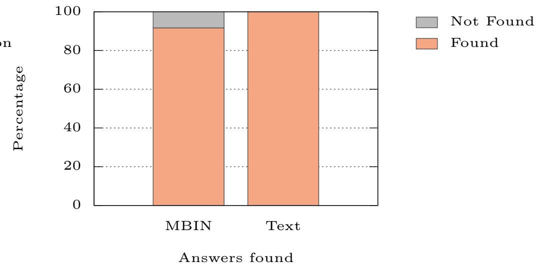
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	20.00	20.00
Mean	29.97	33.33
Standard Deviation	12.00	15.10
Skew	0.37	0.56
Skew Standard Error	0.51	0.51
Kurtosis	-0.54	-0.73
Kurtosis Standard Error	0.99	0.99
Shapiro-Wilk	0.95	0.93
- p-value	0.38	0.15
Significance		
T-Test		-1.00
- p-value		0.33

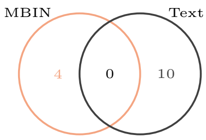
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments
- '*Easy to find what someone has been working on.*' - '*The heat map is great for finding who worked on what.*'

(f) Selection of relevant questionnaire comments

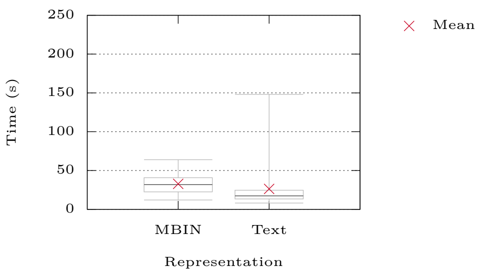
Figure 9.8: Results from Task 8

9.3.9 Task 9: Identify New Knowledge from a Seemingly Unconnected Source; Report Manufacture Issue

See Table 9.10 for the questions relating to this task. Figure 9.9 shows the results for Task 9. Figure 9.9a and Table 9.9b show a significant improvement of approximately 6 seconds in the text-based interface over the model-based interface. This significance calculation is presented in Table 9.9b, the model-based results are normally distributed and the text-based are not, hence a Wilcoxon test being used to determine the significance of the 6 seconds difference in the total times. Figure 9.9c shows the distribution of activity with the difference in completion time meaning no direct comparison can be made between the two interfaces. Figure 9.9d shows how participants were able to answer all questions using the text-based system and nearly all questions using the model-based system. The Venn diagram (Figure 9.9e, shows participants opened over twice the number of reports using the model-based interface compared to the text-based interface.

Questions	
a	Who would you contact for advice about carbon fibre wheel rims?
b	Who would you contact for advice about front inboard suspension?
c	Who would you contact for advice about rear outboard suspension? (UWE only)

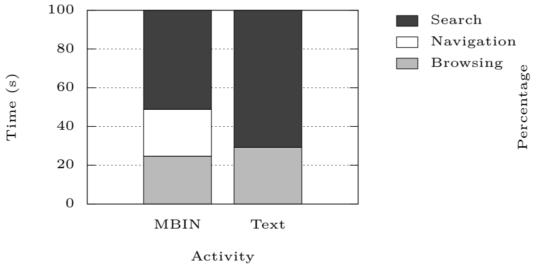
Table 9.10: The questions from Task 9



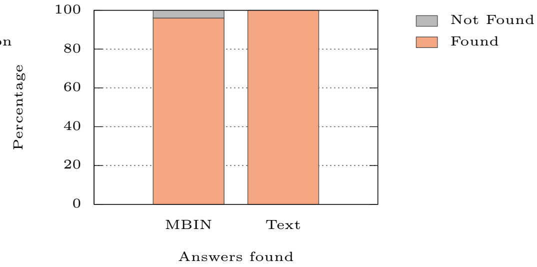
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	20.00	20.00
Mean	32.74	26.21
Standard Deviation	13.64	29.59
Skew	0.57	3.47
Skew Standard Error	0.51	0.51
Kurtosis	-0.48	11.46
Kurtosis Standard Error	0.99	0.99
Shapiro-Wilk	0.96	0.51
- p-value	0.51	0.00
Significance		
Wilcoxon		45.0
- p-value		0.03

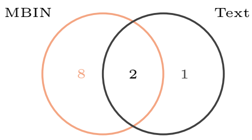
(b) Statistics for the Total Time for question completion for both interfaces



(c) Percentage time spent on search, navigation and browsing for both interfaces



(d) Percentage success rate in ability to answer questions for both interfaces



(e) Venn diagram of the total number of reports opened in each interface

Relevant Participant Questionnaire Comments

(f) Selection of relevant questionnaire comments

Figure 9.9: Results from Task 9

9.3.10 Tasks Summary

The results so far have been presented on a question by question basis, this subsection now examines results from across the nine tasks and relates these back to the Airbus activities Identified in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**. See Table 9.11 for a list of the activities and their ID. Table 9.14 shows the four measures of Total Time, Answers Found, Reports Opened and Activity against each Task and the corresponding activity ID. These are followed by whether there is an increase ($\uparrow$), decrease ($\downarrow$), no difference ($-$) in the model-based interface when compared to the text-based interface, or whether the measures are comparable ($\times$). Improvements in performance are coloured green and degradation in performance coloured red.

Table 9.14 shows how, for the Total Time, the model-based interface was twice quicker than the text-based interface, twice slower and there was no significant difference for the remaining five tasks. For the Answers Found measure, using model-based interface, participants were less likely to find an answer in seven of the nine tasks, more likely in one of the nine tasks and with no difference in one of the nine tasks. The Reports Opened results show that the participants opened more reports on three of the nine tasks, fewer reports on five of the nine tasks with no difference for one of the nine tasks.

In terms of Activity, the breakdown is spilt into time spent searching and browsing. There was no time spent navigating using the text-based interface and so no comparison can be made. For time spent searching, for four of the five tasks, there was less time spent searching using the model-based interface compared to the text-based interface. For time spent browsing, there was a reduction in time browsing in two of the five tasks. At no point was there an increase in either searching or browsing using the model-based interface compared to the text-based interface.

Examining these in terms of Airbus Wing In-Service activities. There is an increase in time for the *analyse extend of damage*, *report manufacture issue*, and *identify new knowledge from a seemingly unconnected past case* activities when comparing the model-based interface against the text-based interface. There is also a decrease in time for the *provide technical drawing*, *instruct on inspection methods*, *provide repair instructions*, *provide aerodynamics analysis*, and *search for similar incident* activities when comparing the model-based interface to the text-based interface. In terms of total time, there was no significance for the remaining activities. To summaries, in there is a decrease in time for five of the 15 activities, and increase in three of the 15 activities and no difference in seven of the 15 activities when comparing the model-based interface to the text-based interface.

When examining the success rate of the participants finding answers to questions, participants were less likely to find an answer using the model-based interface in all but three activities. Participants were more likely to find an answer when performing the *provide information* and *approval for use of new material in repair* activities and there was no difference with the *order item* activity.

Using the model-based interface, participants opened fewer reports for eight of the 15 activities and more for five of the 15 activities compared to the text-based interface. The number of reports opened were the same for two of the activities. Finally, in terms of the activity, participants spent less time searching in six of the 15 activities and the same time in one activity. Time spent browsing was down for three activities and the same for three activities. There was only on occasion where both searching and browsing time was lower. These are all when comparing the model-based interface to the text-based interface. for all those comparable activities, the navigation time is greater (it was not possible to navigate using the text-based interface).

Table 9.13 shows the top four activities with the highest frequency from the 240 reports analysed in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**. These four activities correspond to approximately 95% of all the activities classified. Combining these with the results in Table 9.14 it is shown that the activity with the highest frequency had a reduction in total time, an increase in the answers found and a decrease in the number of reports opened. The second ad fourth highest activity showed no difference in total time, and a reduction the number of answers found and reports opened. The third highest shows an increase in time a reduction in the answers found and an increase in the reports opened. Again these are all using the model-based interface compared to the text-based interface.

ID	Activity
1	analyse extend of damage
2	evaluate proposed repair
3	instruct on inspection methods
4	approval for use of new material in repair
5	order item
6	provide technical drawing
7	report manufacture issue
8	approve repair
9	instruct on inspection methods
10	provide aerodynamics analysis
11	provide information
12	search for similar incident
13	identify most common repair case locations over set period of time
14	identify most common repair case types over set period of time
15	identify new knowledge from a seemingly unconnected past case

Table 9.11: A summary of the engineering activities contained within 240 repair queries received by the Airbus Wing In-Service function during 2013

Task	Activity ID	Total Time	Answer Found	Reports Opened	Activity: Searching	Activity: Browsing
1	1	↑	↓	↑	×	×
2	2, 8	–	↓	↓	↓	–
3	13	–	↓	↓	↓	–
4	5	–	–	↓	–	↓
5	11, 4	–	↓	↑	↓	↓
6	6, 9, 3	↓	↑	↓	×	×
7	10, 12	↓	↓	–	×	×
8	14	–	↓	↓	↓	–
9	15, 7	↑	↓	↑	×	×

↑	Increase compared to text-based interface
↓	Decrease compared to text-based interface
–	No difference
×	Not Comparable
Red	Degradation
Green	Improvement

Table 9.12: A comparison the model-based interface against text-based interface measures across tasks

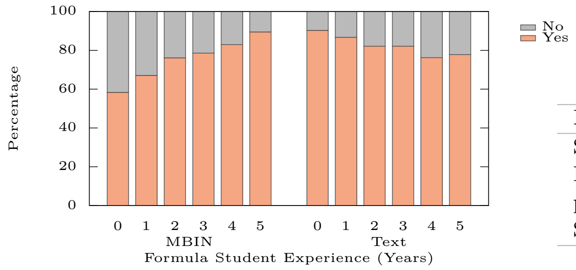
ID	Activity	Frequency
3	provide repair instructions	118
2	evaluate proposed repair	117
1	analyse extend of damage	89
8	approve repair	38

Table 9.13: A summary of the four highest frequency engineering activities contained within 240 repair queries received by the Airbus Wing In-Service function during 2013

9.3.11 Demographics

The results shown so far in this section have been broken down on a question by question basis. In this subsection, the results are considered across the nine pairs of questions with the goal of determining which, if any of the demographic data, captured in the questionnaire, has an effect on the time to complete and success rates. In total, CAD experience, general automotive experience, age, Formula Student experience, leadership role and University where examined. In terms of time to complete each task, dividing the participants by these categories left sample sizes that were too small to determine any meaningful and reliable results. The success rate across all tasks however did show some interesting findings in regards to the subsets of Formula Student experience and whether the participant was in a leadership role or not.

Figure 9.10a shows the participants' success rate in answering all questions broken down by their experience in Formula Student from zero years to five years. The graph shows a linear correlation between the greater the experience in Formula Student, the more likely the participant will find an answer using the model-based interface. At zero years, on average participants are able to find a result approximately 60% of the time and this steadily rises to nearly 90% at five years. Equalling the maximum level achieved by participants using the text-based system. In contrast, the results for the text-based system start at the maximum of nearly 90% before dropping to just below 80% for the most experienced participants.



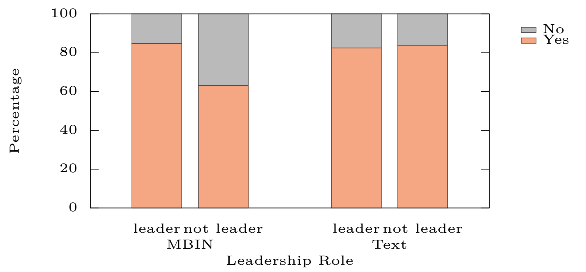
(a)

Measure	MBIN	Text
Slope	5.88	-2.70
R-squared	0.97	0.90
p-value	0.00	0.00
Standard Error	0.55	0.45

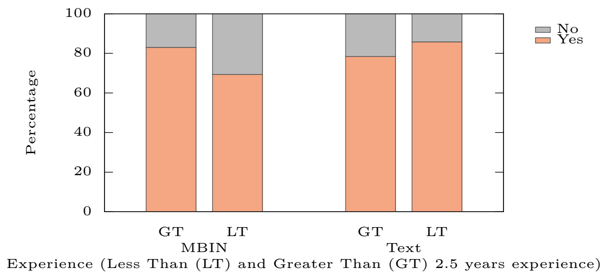
(b)

Figure 9.10: Answer success rate by Formula Student experience and associated linear correlation statistics for both interfaces

The second finding relates to the question answering success rate and whether the participant is in a leadership role or not. Figure 9.11a shows the mean success rate broken down by whether the participant is in a leadership role, for both the model-based and text-based interface. The graph shows a success rate of over 80% for all those participants apart from those in a non-leadership role using the model-based interface where the success rate falls to approximately 60%. This aligns somewhat with the results shown in Figure 9.10a given, the more experienced a participant is, the more likely they are too be in a leadership role. Figure 9.11b shows the results for participants split by their Formula Student experience (above (GT) and below (LT) 2.5 years). Comparing Figures 9.11a and 9.11b directly shows the difference between the effect of leadership against experience. The most obvious difference between the two is that those with less experience are more likely to find information with the model-based interface than those who are not leaders. Given some participants are both leaders and have less experience, this indicates that both leadership and experience have an influence on the ability to find an answer using the model-based interface.



(a) Leadership



(b) Experience

Figure 9.11: Answer success rate by participant leadership role (a) or experience (b) for both interfaces

Exploring the effect of Formula Student experience further, Table 9.11 shows the summarised results for: total time, answers found, reports opened, and searching and browsing activity, split by above and below 2.5 years experience. Appendix A.4 shows the full set of results. For total time, the lesser experienced group showed four significant differences in time, with two increased and two decreases in time. For the more experienced group, there is only one significant improvement in time. The differences in answers found between the two interfaces is identical for both groups. For the reports opened however, the more experienced group of participants opened fewer reports using the model-based interface compared to the text-based interface in seven of the nine tasks compared to four of nine to the least experienced group. The results for searching and browsing show a similar picture to Table 9.14 and the results for the question level summary - where comparable, the participants spend less percentage time performing these activities using the model-based interface over the text-based interface, this is understandable given within the model-based interface the participants are also engaged in navigating the virtual environment. The one exception being the more experienced group and the searching activity for Task 2. Here the experienced group spent longer searching that using the model-based interface than the text-based interface. There does not seem to be an obvious reason why this would be the case.

		Total Time		Answer Found		Reports Opened		Activity: Searching		Activity: Browsing	
Task	Activity ID	<2.5	>2.5	<2.5	>2.5	<2.5	>2.5	<2.5	>2.5	<2.5	>2.5
1	1	↓	–	↓	↓	↑	↓	×	–	×	↓
2	2, 8	–	–	↓	↓	↓	↓	↓	↑	↓	↓
3	13	–	–	↓	↓	↓	↓	↓	–	↓	–
4	5	–	–	↑	↑	–	↓	↓	↓	–	↓
5	11, 4	–	–	↓	↓	–	–	–	↓	↓	–
6	6, 9, 3	↑	↑	↑	↑	↓	↓	×	×	×	×
7	10, 12	↑	–	↓	↓	↑	↓	×	↓	×	↓
8	14	–	–	↓	↓	↓	↓	↓	–	↓	–
9	15, 7	↓	–	↓	↓	↑	↑	×	↓	×	↓

Table 9.14: A comparison the model-based interface against text-based interface measures across tasks for two groups with above (>2.5) and below (<2.5) years Formula Student experience

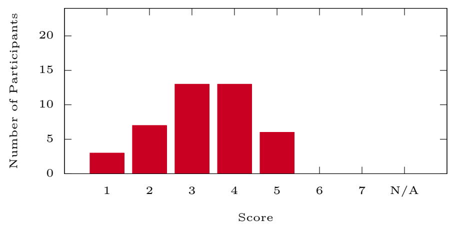
↑	Increase compared to text-based interface
↓	Decrease compared to text-based interface
–	No difference
×	Not Comparable
Red	Degradation
Green	Improvement

9.3.12 System Usability Questionnaire

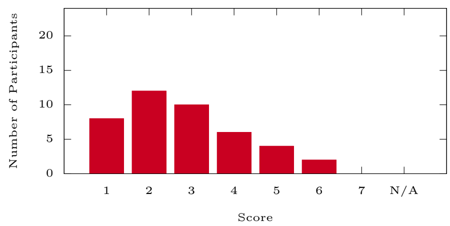
This section presents the participant responses for the System Usability Questionnaire (See Appendix A.2). The scale runs from 1 (strongly agree) to 7 (strongly disagree) with a not applicable (N/A) option. Figure 9.12 show the responses to each question in turn with the question is included as the individual figure caption. Figures 9.12m and 9.12n show the combined results for questions 1 to 8 (system use) and questions 9-12 (interface quality) - as recommended within the IBM guidelines for the questionnaire [79].

Looking at Figures 9.12a to 9.12l, the results are relatively mixed, however, if the score of 4 is taken as a mid-point and the results spilt from positive (score of 1, 2, or 3) and negative (a score of 5, 6, or 7), then across all questions, the model-based interface scores positively. This is more evident in the responses to some questions over others, Figure 9.12g for example shows how the participants felt the system was easy to learn. Compare this to Figure 9.12e for example, where overall the participants felt they were able to efficiently complete the tasks, however, the number of participants scoring 5 and 6 show that some participants felt the opposite.

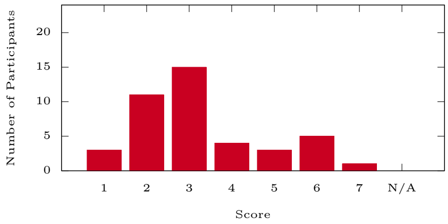
Figure 9.12m shows the accumulated responses from questions 1 to 8 that are used to evaluate the system usage. These results show a clear positive response to using the system, with there being a peak at the score of 2. Figure 9.12n shows the accumulated response from questions 9 to 12 that are used to evaluate the quality of the interface. Here, there is a peak at the score of 3, however, most scores still sit on the positive side of the scale. This is however an indication that there may have been some issues with the quality of the system.



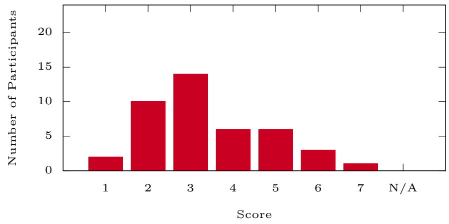
(a) (1) Overall, I am satisfied with how easy it is to use this system.



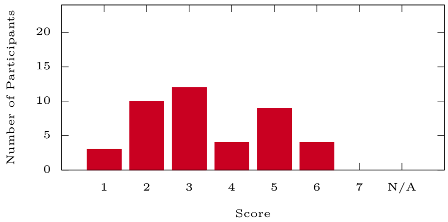
(b) (2) It was simple to use this system.



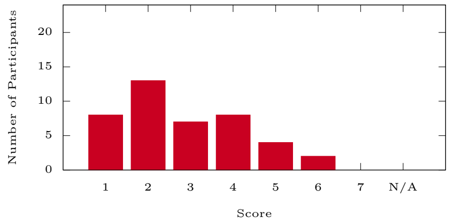
(c) (3) I could effectively complete the tasks and scenarios using this system.



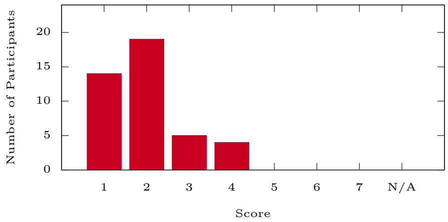
(d) (4) I was able to complete the tasks and scenarios quickly using this system.



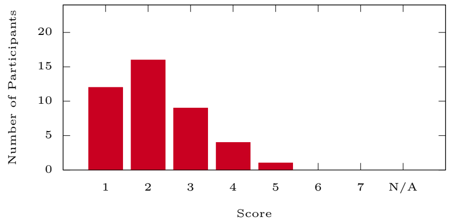
(e) (5) I was able to efficiently complete the tasks and scenarios using this system.



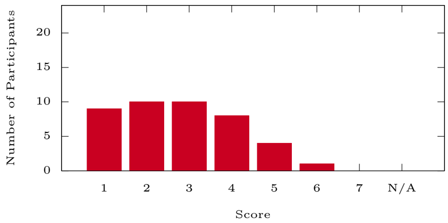
(f) (6) I felt comfortable using this system.



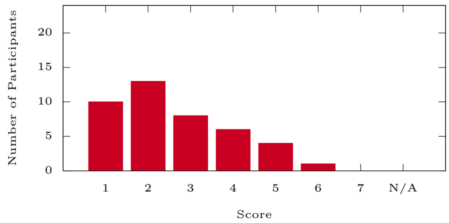
(g) (7) It was easy to learn to use this system.



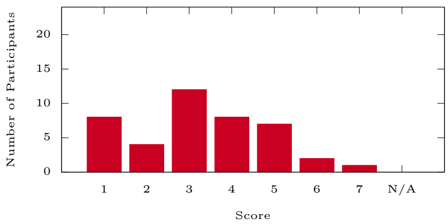
(h) (8) I believe I could become productive quickly using this system.



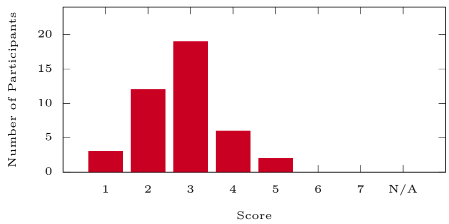
(i) (9) The interface of this system was pleasant.



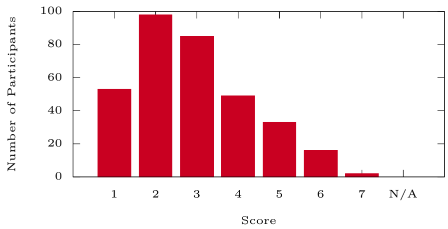
(j) (10) I liked using the interface of this system.



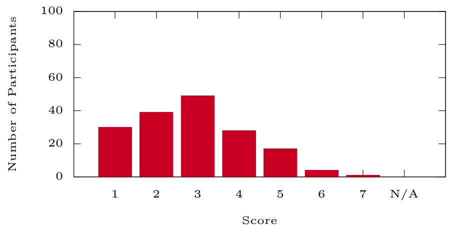
(k) (11) This system has all the functions and capabilities I expect it to have.



(l) (12) Overall, I am satisfied with this system.



(m) An accumulation of the scores from question 1 to 8 that form the system usability proportion of the questionnaire.



(n) An accumulation of the scores from question 9 to 12 that form the system interface quality proportion of the questionnaire.

Figure 9.12: The responses to the System Usability Questionnaire

9.3.13 Participant Comments

When exploring the more general questionnaire comments, a number of themes can be established. These are shown in Table 9.16 and include Technical Problems, Search Problems, Text More Familiar, Need More Time, Preference and Both at Once. Table 9.15 shows a list of the comment themes and their descriptions. The 'themes' were derived naturally found within the comments, inline with the procedure for thematic analysis as outlined by Moria and Bird [125].

ID	Category	Description
1	Technical Problems	Comments relating to system technical problems, the system crashing for example
2	Search Problems	Comments relating to the search engine not performing as expected
3	Text More Familiar	Comments that the text-based interface was more familiar than the model-based interface
4	Need More Time	Comments that more time was needed to become familiar with the model-based system
5	Preference	Comments stating a preference for one interface other the other
6	Both at Once	Comments stating that the participant would prefer access to both interfaces at the same time

Table 9.15: A list of the participant comment categories and their description

Firstly, one can not ignore that some users had technical problems (ID 1 in Table 9.15). The system re-set if the user pressed the back-button or refreshed the page and most if not all of the reported 'crashes' were due to the participant accidentally performing one of these actions. Other issues, such as text-boxes not clearing or the search boxes not responding to numbers, were consistent between the two interfaces and as such should not have had an effect on the results presented.

Secondly and categorised as 'Search Problems', the ranking method employed, the TD-IDF weighting (see **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**), was not always the most useful measure to rank by. This speaks to the wider difficulties in developing successful enterprise-type search engines. The ranking method was the same for both interfaces and as such this was mitigated against in the design.

The 'Text More Familiar' category shows some participants noted how the text-based interface was more familiar (from life outside the study) and as such may have had an advantage, this was reflected in comments that stated the participant felt they would improve with more time with the system (the 'Need More Time' category). The comments seem to suggest that the 15-20 minutes provided at the beginning of the study was adequate for some people but not all. It is not possible to say how much of an effect that this may have had on the results presented. However, the finding that those more experienced in Formula Student are more likely to successfully answer questions does suggest that a longer period of familiarisation would improve results. Repeating the study after providing the system to participants over periods of months should mitigate this.

In terms of the user preferences ('Preference' in Table 9.15) between the two interfaces there is a far larger number of comments in favour of the model-based interface than are both critical of the model-based interface or in favour of the text-based interface. Participants both understood the potential benefits of the model-based interface in spite of any short-comings in the implementation or search engine. A number of comments note how the system was intuitive to use, with some noting how the model-based interface made searching for information easier. The final category and relating to the preference, 'Both at Once', show some participants would have preferred to have access to both interfaces at the same time and be free to switch between the two.

ID	Category	Participant Questionnaire Comments
1	Technical Problem	- *'pan speed when zoomed in is too fast'* - *'I think this could be a valuable tool with some refinement and sorting teething problems.'* - *'computer crashed using the system'* - *'the system needs a few modifications.'* - *'Improve the search system by including metadata'* - *'spaces in the beginning of the search box are not ignored as well.'* - *'Right hand search box didn't seem to work with numbers?'* - *'Search boxes need to be cleared between windows'* - *'It could be a bit more refined.* - *'Noticed capitals change what is found in a search.'* - *'Search engine seems clunky.'*
2	Search Problems	- *The relevance of the research papers with respect to the results was not necessarily in the best order at times.'* - *'Search requires some refining.'* - *'Very few clicks needed! Search could be refined.'* - *'If search was a little more intelligent (showed more, refine by document type).'* - *'Provided the search function has more filtering'*
3	Text more familiar	- *Text based is intuitive because I'm used to it, but 3D search was also very intuitive.'* - *'Likely because it's new. I've used textbased systems my whole life.'* - *'I use search engines more frequently than CAD.'* - *'Yes, but much more used to text based.'*
4	Need more time	- *'Needed some getting used to.'* - *'I believe I would get better with time.'* - *'The route to find things was often not straight forward'* - *'I think it could get comfortable with time'* - *'needs a little more practice.'* - *'Unsure. Possibly with practice.'*
5	Preference	- *'In general it is a great idea.* - *'Lovely! Easy to access part information, find out part names/details.'* - *'Easy to find what someone has been working on.'* - *'Found car representation more intuitive but it was limiting at times'* - *'Made searching for areas much easier'* - *'Was very intuitive'* - *'Very simple!'* - *'Definitely easier to use the visual system.'* - *'Made the search for information a lot easier.'* - *'The route to find things was often not straight forward'* - *'Found it was easy to find the require files'* - *'Very few clicks needed! Search could be refined.'* - *'Made searching by area very easy'* - *'Felt easy and comfortable to use'* - *'Made a lot of sense, felt natural.'* - *'Absolutely! Easy and intuitive navigation.'* - *'system was fairly simple in nature'* - *'Very intuitive.'* - *'felt fully in control after only a few minutes'* - *'Incredibly intuitive'* - *'Fairly self-explanatory'* - *'It is very easy to get the hang of.'* - *'If some amendments were made, it would be a great aid.'* - *'Could be very useful for searching in this way, especially looking through teams I don't know'* - *'Just as long as a traditional search engine was also available!'* - *'Definitely more efficient for finding information on particular components'* - *'In general it is a great idea.'* - *'This system makes so much sense than conventional system. 10/10. Any chance of us using it anytime soon?'*

		- *'It was really good much better than normal folder structures or my experience of Siemens TeamCentre.'* - *'It was easier to be general with the text based search engine'* - *'Usually, some searches (e.g. rules) are easier in text based.'* - *'The 3D system was confusing'* - *'Only for certain applications, text is fine on a daily basis.'* - *'It helps me understand in this visual 3D interface'* - *'I think for CAD models a 3D model is more useful and helps me understand, however the 3D model is often confusing when you have to search something that doesn't have a specific part'* - *'The 3D model just adds another step as to use the heat map you must use text anyway and if clicking does not give what you want you revert back to text based system.'* - *'The 3D model search is best when needing to find a document to a known part. The text based search is best for names and specific documents that might relate to more than one part. '*
6	Both at once	- *'Using both the text search and 3D model in conjunction would have been preferred.'* - *'It was quite easy to use and would have been better if the answers to questions would have been found without being restricted to one method.'* - *'The ability to switch between the car and the search screen would have been helpful at times.'* - *'Some were very easy others very hard: having both available would be useful'* - *'would have been more efficient if both tools were available'* - *'Just as long as a traditional search engine was also available!'* - *'If it also had the traditional interface as well!'* - *'Would be good to include a standard text search as a back up system.'* - *'They are both more efficient depending on the given task.'* - *'Both can be useful.'* - *'I'd prefer to have both available'* - *'Both searches provided results that I understood.'* - *'The 3D model search is best when needing to find a document to a known part. The text based search is best for names and specific documents that might relate to more than one part. '* - *'The efficiency depended greatly on what was being asked. If there was a question about a specific person, 3D model was useful. Anything more general, the search engine was useful.'*

Table 9.16: A selection of categorised participant comments from the questionnaire

9.3.14 Seven Characteristics of Information

Chapter 3: Literature Review introduce the seven characteristics of information, see Table 9.17 for a summary of the definitions. This subsection now explores the study results within the context of these seven characteristics to understand the effect of a model-based interface.

Character	Description
Accessibility	The users' ability to access a piece of information. I.e. '*I know the information is in the system but how do I find it?*'
Usability	The ease of use of the system to support the access of information.
Currency	The time taken to access a piece of information.
Context	The relationship between information and the information need i.e. an engineer searching for the term 'cat' is more likely to be searching for a catalytic converter rather than the animal.
Accuracy	How well a piece of information meets an information need. I.e. '*I want to find a picture of a cat but I keep getting pictures of stuffed toys, the search is not that accurate*'.
Availability	Whether a piece of information is available for access or not. I.e. '*I know the information exists but the system isn't able to access it*'.
Relevance	A measure of appropriateness of a piece of information to an information need, i.e. '*I searched for cat and got a picture of a dog. Dogs aren't relevant*'.

Table 9.17: Seven characteristics of information and the corresponding descriptions used in this thesis

Accessibility

Accessibility relates to the users' ability to find a piece of information and there are a number of factors associated to the study that effected how accessible information was. Table 9.14 shows how users were less likely to find results in seven of the nine tasks, more likely in one and the same in one. Based on these findings, the model-based approach makes information less accessible, however, Table 9.14 also shows that participants opened fewer reports and spent less time searching and browsing using the model-based approach. This could be an area that uses the advantage of the document classification-based approach to indexing (as outlined in **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**), given an improvement in precision and a reduction in recall will mean uses will see fewer but more 'correct' results.

Usability

The usability is the ease with which users use the system. The results relating most to the usability of the system are those shown in Figure 9.12m - the system usability score. The figure shows participants agree that the system is easy and intuitive to use when compared to the text-based system.

Currency

The currency within the context of this thesis is the time taken to find information. Figure 9.14: shows that for total time, the model-based interface was twice quicker, twice slower and the same compared to the text-based interface. The effect on a model-based interface on time is then a mixed one.

Context

Context is the relationship between the information and the information need. Part of the justification for the model-based interface was the three-dimensional representation of the product would provide the real-world context of the physical artefact. This is reflected in the results relating to context are shown in Table 9.16. These are some examples:

> - *'It helps me understand in this visual 3D interface'*
>
> - *'I think for CAD models a 3D model is more useful and helps me understand, however the 3D model is often confusing when you have to search something that doesn't have a specific part'*

The second comment here is an indicator that the participant did find the context of the three-dimensional model confusing when looking for information that did not have an obvious place within the model.

Accuracy

Accuracy is a measure of how well a piece of information satisfies and information need. The following comments are two examples relating to accuracy and taken from Table 9.16.

> - *'The 3D model search is best when needing to find a document to a known part. The text based search is best for names and specific documents that might relate to more than one part. '*
>
> - *'The efficiency depended greatly on what was being asked. If there was a question about a specific person, 3D model was useful. Anything more general, the search engine was useful.'*

These two comments highlight where participants show a preference for the the model-based interface or the text-based interface. They show users find the of the model-based interface improves accuracy for certain pieces of information over others: part and people specific information in particular.

Availability

Availability is whether information is available for access or not. Given the index behind the two interfaces was the same, the vast majority of documents were treated in exactly the same way. One area where they differed was in relation to Task 6: provide technical drawing; instruct on assembly method; provide design instruction; The text-based index could not find the drawing while the model-based could. This then is a benefit of a model-based over the text-based interface.

Relevance

Relevance is a measure of the appropriateness of a piece of information to meet the information need. These are some participant comments taken from Table 9.16.

> - *The relevance of the research papers with respect to the results was not necessarily in the best order at times.'*
>
> - *'Search requires some refining.'*
>
> - *'Provided the search function has more filtering'*

Given the indexing approach was the same for both model-based and text-based interface, these results reflect how the indexing method itself could be improved.

Summary

Over the seven characteristics of information then, it can be said that the model-based interface had an effect of six of the seven characteristics. With the design of the system/study resulting in no difference found in the relevance characteristic (both interfaces used the same index hence, relevance was the same). Of the six characteristics where an effect was found, the positive effects generally related to very specific use cases (searching for part related information for example), and the negative affects also meant the text-based system was preferred in some specific cases (general queries, or those not relating to specific areas of the car for examples). The general picture is then mixed, however, the model-based interface does show some promise in some areas.

9.4 Discussion and Conclusion

This section explores the general trends visible in the results before reflecting on the research approach taken. Finally, the affordance of model-based information navigation are highlighted in answer to the fourth and final research question on whether model-based information navigation improves information access and knowledge discovery.

9.4.1 General Trends

Overall the results are mixed if one is looking for a clear improvement in information access and knowledge discovery. Exploring the key measure of time to complete each task first, over the nine pairs of questions four showed a significant difference: two in favour of the model-based interface and two in favour of the text-based interface. For the remaining five questions there was no significant difference. This is an interesting finding given that the model-based interface adds an additional and complex step to the search process (navigating the three-dimensional model) and yet in most cases, this does not slow the search process. Where the participant activity can be broken down into search, navigation and browsing, and the results allow a direct comparison of time spent performing each task, they show the time spent navigating reduces the time spent formulating search queries and occasionally some time spent browsing.

The two largest differences in times, Tasks 1 (approximately 20 seconds in favour of the text-based interface) and Task 6 (approximately 35 seconds in favour of the model-based interface), both had additional factors influencing the search. In Task 1, the fact that one document is related to a specific component and the other is a more general document highlighted how some users struggled with the concept of where that document belongs in the model and this could have had a negative impact on the model-based timing which would not have been present in the text-based timing. With Task 6, the text-based method is not capable of indexing drawings whereas the model-based interface did. The number of reports opened via the text-based interface were also far higher with this question than any other. The 35-second difference is then clearly down to the fact that users could not find the actual drawing using the text-based interface. While the validity of the time difference may not then substantial, the fact that the model-based interface allows for the indexing and finding of drawings has been highlighted as an affordance of the model-based interface.

The participants' ability to find answers to questions was lower with the model-based interface in seven of the nine questions, roughly the same in one question and higher in the aforementioned drawing question. The results do however show that certain factors, Formula Student experience and being in a leadership role, does improve this with both interfaces performing roughly equally by five years experience or being in a leadership role. The description of those graphs (Figure 9.10 and 9.11a) shows how other demographics were explored with no significant findings. It is interesting that CAD experience had no effect, given the users are effectively interacting with the CAD model to find information. This could be due to the system being web-based (and not a CAD package add-on) and as such, all users were required to learn a new set of interaction techniques and that those techniques, and the model-concept itself, were deliberately designed to be simple and intuitive whereas CAD packages can be more complex to interact with. There was also no correlation found between likelihood of answering questions and general automotive experience. This is another interesting finding however we cannot be sure of this as there were only a small number of participants who had automotive experience outside of university.

As previously stated, the results revealed that experience in the Formula Student project and whether the participant is in a leadership role impacts performance. This is likely due to the participants' familiarity with the model they see on-screen and the understanding of the wider project - where parts are on the car. Returning to the Airbus use-case, the In-Service department design repairs for in-service products and as such, the engineers working in those departments are familiar with the product and hence, should perform better than the students used in this study. It can be said however that the results show that the model-based interface is more suitable to products after the design phase of the product life-cycle, where engineers are more familiar with the product.

When combining the results with the list of Airbus Wing In-Service activities presented in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design** the picture is mixed. If one considers an improvement in information access and knowledge discovery as a reduction in total time, an increase in the number of answers found and a reduction in the number of reports opened, then the results show an improvement for the most frequent activity (*provide repair instruction*). However, the second and fourth most frequent activities show no difference in total time, a reduction in answers found and a reduction in the number of reports opened and the third highest activity shows an increase in time, a reduction in answers found and an increase in the number of reports opened. There is then no clear "better" interface when considering these measures. The model-based interface is shown to performed "better", "worse", and the same as the text-based interface.

There were a number of comments that describe the model-based interface as intuitive, simple and easy to use and that it aided and supported the finding of information. This was reflected in the responses to the scores submitted in response to the system usage. Alongside this, and on a more cautious note, in terms of a day-to-day implementation, participants also commented that they would prefer both systems side by side. While participants enjoy using the system, there is a learning curve associated with both familiarisation with the project as well as the system. Providing both forms on interface mitigates this and the expectations would be that over time, users would naturally migrate to the model-based interface.

9.4.2 Reflections of Method/Approach

Tasks 3, 8 and 9 related to knowledge discovery. Tasks 3 and 8 show no significant differences between the two interfaces while Task 9 found an approximately six second improvement in the text-based interface of the model-based interface. These findings are probably the most surprising considering the first two (3 and 8) could have been answered using the heat map and without opening any reports. In terms of Task 9, on reflection the question could have been formulated in a manner that better reflects the task of identifying information from seemingly unconnected components. Comments relating to needing to completely trust the heat map may be having an influence over these results, given that if this were the case, the participants would follow up the initial heat map query by searching through results. The general sense from comments and post-study discussions with participants is that these forms of search will improve once participants are familiar with the system and have developed their own search 'strategies'. One participant commented on the questionnaire that they used the heat map to narrow their search. Participants familiar with text-based search will have developed these for search and 1) the 15 minutes familiarisation time may not have been long enough to develop strategies for the model-based interface and 2) the knowledge discovery tasks through the model-based interface requires a new interface compared to text. The knowledge discovery findings are then inconclusive and further work study is required that provides participants with more time with the system to learn search strategies.

The final reflection is on the decision to include the UWE participants in these results, given the model-based interface that they used was changed for the participants at the other three universities. As stated in **Chapter 8: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 1: Study Design**, the justification for including these participants was that 1) the effect on performance would be small compared to the effects of the dependent variable (i.e. the interface (model-based and text-based)) and 2) the effect would be largely mitigated by averaging results across all universities. The results should still however be interpreted with this in mind.

9.4.3 Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery

In conclusion, the study and results shown in this chapter suggest that in answer to whether model-based information navigation improves information access and knowledge discovery:

- There is no significant difference in time to complete a search in spite of the addition of a new stage in the search process
- The system structure of the model-based interface allows for non-text based documents to be indexed, making up for the inherent limitations in traditional text-based search
- Participants new to the product will be less likely to find answers than those who are familiar with the project

- Leaders are more likely to find information that those who are not leaders
- Participants enjoy using the model-based interface and find it intuitive, easy and simple to use
- For the seven characteristics of information, the model-based interface shows some improvements in response to specific information needs

This chapter began by extracting the information needs of Airbus In-Service engineers before translating these over to a Formula Student use-case and testing a model-based interface to information navigation and knowledge discovery with a basic implementation of the system. This bare-bones interface was deliberate with an attempt to identify the effect that a three-dimensional visualisation has over a more traditional text-based search engine and visualisation technique. This chapter did not explore a system that incorporated the findings from the other three research questions, and the natural step from here would be to gather and integrate the findings and implement the system complete. The remaining chapter of this thesis, **Chapter 10: Conclusion, Implications for Airbus, and Future Work** explores this through a discussion of the work presented throughout this thesis.

Chapter 10

Conclusion, Implications for Airbus, and Future Work

10.1 Introduction

This chapter closes this thesis by exploring the findings from the four research questions before discussing them within the context of the Airbus Wing In-Service team. The final section of this chapter is then dedicated to outlining a number of topics deemed worthy of future research.

10.2 Contribution to Knowledge

The aim of this thesis was: *The development of a framework for model-based information navigation and knowledge discovery in large engineering organisations* Model-based information navigation being the process of searching for and finding engineering information via a three-dimensional model of the product. This aim has been met through the creation of a methodology for model-based information navigation of engineering information in a large multi-national engineering organisation. The methodology comprises of:

- approaches for indexing engineering documents against model/product structures
- techniques for the navigation of model-based three-dimensional virtual environments

- means of displaying engineering information within model-based three-dimensional virtual environments

The industrial partner in this work, the Airbus Group, provided the data and use-cases that allowed for the design and development of both a test rig system and the means to test the system in a manner that appropriately represents the ‘real-world’ of Airbus operations. The thesis was aligned with the Wing In-Service team in Filton, Bristol, UK. A team of front line engineers who respond to maintenance queries and repair design request from customer airlines. These responses are time critical and as such, the team would benefit from a mean of better information and knowledge dissemination.

The model-based concept was investigated through four research questions that were derived through a scoping study, presented in **Chapter 2: Industrial Context and Scoping Study** of this thesis, and through an iterative prototyping of a test rig, presented in **Chapter 4: Aim, Methodology and Research Questions**. These four questions:

RQ 1 What are the most appropriate techniques for a model-based approach to document indexing?

RQ 2 What are the most appropriate techniques for navigation information within a model-based virtual environment?

RQ 3 What are the most appropriate techniques for displaying information within the model-based virtual environment?

RQ 4 How does model-based information navigation improve engineering information access and knowledge discovery?

Research Questions 1, 2, 3, and 4 are presented in Chapters 5, 6, 7 and 8 and 9 respectfully. This section now explains the findings from each of these questions before the next section describes the implications of the findings to the Airbus Wing In-Service team.

10.2.1 Research Question 1

The first research question considered document indexing and determining appropriate techniques for relating documents to the product structure. Table 10.1 shows the findings that document classification techniques can be used to improve the document indexing technique over the traditional TF-IDF indexing methodology when classification is used to classify documents against product components. These findings answer Research Question 1 and are significant in that they show an engineering specific solution can improve search in a field that literature (**Chapter 3: Literature Review** and **Chapter 5: Appropriate Techniques for a Model-Based Approach to Document Indexing**) finds can be found lacking. The novelty here being the demonstration that document classification against the product structure (components, systems and sub-systems) can be used to improve precision/recall when compared to a more traditional TF-IDF search engine (searching for the names of those components).

Topic	Finding
Model-Based Document Indexing	Document classification techniques can improve document indexing using classification against the product structure.
	Manual indexing against the product structure can improve document retrieval of non-text based documents such as CAD files.

Table 10.1: The findings from Research Question 1

10.2.2 Research Question 2

The second research question considered the user navigating the model-based virtual environment to find information. This question arose from the problem that users of a model-based system will need to find information buried in the depths of a three-dimensional model. Any model-based system must facilitate and support the user in their efforts to access that information if the users it to 1) access information and 2) access information in a manner that is not laborious (otherwise users will not use the system). The answer to the second research question is a list of techniques for interacting with the three-dimensional space of the model-based approach. These ranged from the more traditional exploded views and text product lists to more novel 'onion peeling'/MRI slicing the model. See Table 10.2.

The list of candidate navigation techniques are well-founded in literature and from similar fields that use three-dimensional user interfaces. Onion peeling being inspired by medical MRI slicing being the most obvious example. Where the drag, product tree list, onion peeling, exploded view, zoom, and pan are all aimed at the manipulation of components, the heat map technique was designed as a means of knowledge discovery, showing the location and frequency of search terms across the product structure. The list is not exhaustive and would benefit from further research, however, the findings show that together they are sufficient for the implementation of a model-based interface.

Topic	Finding
Model-Based Navigation	A rage of techniques for the navigation of engineering model-based virtual environments. These are: heat maps, product tree list/textual list, onion peeling/MRI slicing, drag/manual manipulation of components, reset mesh, exploded view, zoom, and pan.

Table 10.2: The findings from Research Question 2

10.2.3 Research Question 3

The third research question related to displaying information within the model-based environment - essentially, what should information 'look like'. Answering the question involved a review of similar fields, ranging from road signs to engineering drawings. The output being a list of visual information objects, markers (including the product components themselves) that inhabit the model-based environment and have information associated to them and act as a gateway to that information. Two-dimensional virtual environments, online maps (Google Maps) for example, use markers in this means to access information such as business details or sites of historical interest and photographs. Prior to the work presented here, this had not been applied to engineering three-dimensional virtual environments.

The contribution to knowledge resulting from Research Question 3 are the visual information objects: component-of-interest, section-of-interest, feature-of-interest, and point-of-interest. A range of markers designed for the identification of information in engineering three-dimensional virtual environments. The relevance of this contribution stretched beyond search and encompasses virtual and augmented reality applications, where their growth in population is also leading to their use for the display of information. See Table 10.3 for a summary of the findings.

Topic	Finding
Model-Based Information Representation	A range of markers for the identification of information within engineering model-based virtual environments. These are: component-of-interest, section-of-interest, feature-of-interest, and point-of-interest
	Participants would prefer the use of components-of-interest for the display of information in model-based virtual environments.
	Region-of-interests, sections-of-interests, and points-of-interest are also required in certain scenarios.
	There is some ambiguity between the use of directional and non-directional points-of-interest.

Table 10.3: The findings from Research Question 3

10.2.4 Research Question 4

The fourth and final research question explored the actual process of finding information within a model-based environment using a direct comparison between it and a more traditional text-based approach to finding information. The study used a number of engineering tasks based on the Airbus Wing In-Service operations and mapped over to Formula Student surrogates. The Formula Student surrogate participants completed the tasks while the system logged their activities. A questionnaire based on the IBM CSUQ questionnaire was also used to capture participant demographics and their perception of using the model-based interface.

There were a number of findings associated to this study, the first being that there was little difference in the time taken to find information between the model-based and text-based approaches. With five of the nine tasks showing no significant difference in time, two showing a sorter time and two showing a longer time between completing the tasks with the model-based interface compared to the text-based interface. This find is significant given the model-based approach adds a complex step (navigating the three-dimensional model) to the search process and yet the total search time is not significantly slower.

The second finding is that for documents such as engineering drawings, traditional indexing techniques are not able to extract text from the drawings and as such, are not able to index the files. Meaning they can not be found using a traditional search engine. The model-based approach used to index in the study, mitigated this and the result was a search system that indexed engineering drawings alongside text documents.

The third finding discovered that users inexperienced in the product/project are less likely than those who are experienced or in a leadership role to find information using the model-based approach than a text-based approach. The least experienced participants found answers approximately 60% of the time compared to approximately 90% for the most experienced/leaders. 90% being approximately the maximum success rate seen for the text-based interface. This shows that the model-based interface can perform as well as the text-based interfaced for experienced users/leaders.

The fifth finding showed that users enjoyed using the model-based approach and found it intuitive to use. These findings are repeated in Table 10.4. These findings combine to show a system that users inherently understand, enjoy and once experienced in the product/project, will benefit from the approach to finding information and discovering knowledge.

The final finding showed that, depending on the information need of the participant, some improvement could be seen in the seven characteristics of information.

Topic	Finding
Model-Based Information Navigation	There is no significant difference in time to complete a search in spite of the addition of a new stage in the search process.
	The system structure of the model-based approach allows for non-text based documents to be indexed, making up for the inherent limitations in traditional text-based search.
	Participants new to the product will be less likely to find answers than those who are familiar with the project.
	Leaders are more likely to find information that those who are not leaders.
	Participants enjoy using the model-based approach and find it intuitive, easy and simple to use.
	For the seven characteristics of information, the model-based interface shows some improvements in response to specific information needs

Table 10.4: The findings from Research Question 4

So far in this thesis, the findings have been discussed in terms of the research questions, this chapter now continues by discussing them in the wider context of the Airbus Group Wing In-Service team.

10.3 Implications for the Airbus Wing In-Service Team

This thesis began with an introduction to the Airbus Group and the Wing In-Service team and as such it is fitting that the findings of this thesis now be discussed within the context of Airbus. This section does this by running through the research questions once again however, this time with the In-Service group in mind.

10.3.1 Research Question 1

In terms of document indexing, as discussed in **Chapter 2: Industrial Context and Scoping Study**, the Wing In-Service team have implemented a content-based search engine - Daedalus. Daedalus uses the raw text from the maintenance and repair reports alongside an ontology to expand search queries and return a more intelligent set of results. The limitation of this process is however the same as all text-based search engines in that media such as drawings, photographs, CAD files, CFD/FEA analysis are not text based and as such are not able to be indexed. The ability to link directly to a CAD file or a set of analyse without having to traverse a report first, would reduce the time to find that document and ultimately reduce the time to provide a repair.

The fact that the model-based approach to indexing demonstrated that these media could be indexed alongside text documents could result in a far richer and more useful set of results. In addition to this, the indexing findings showed that document classification can be used to index documents against the product structure. This means that the indexing process for these additional formats could be automated and remove the manual effort involved in indexing via the means used in the Research Question 4 study. Ultimately, there is an opportunity to improve search beyond that of the existing system as well as introduce mixed media search, both of which would have a beneficial impact on business performance.

The final implication for Airbus relates to issues with synonyms and documents written in different languages (synonyms being more than one term/part number for the same thing). The product is the same regardless of what components are being called, and as such, all documents, regardless of the language that they are written in, are able to be found.

10.3.2 Research Question 2

Research Question 2 generated a set of techniques for navigating the three-dimensional model-based interface. Simply put, the findings from Research Question 2 should allow for the implementation of a usable model-based interface as they cover a complete set of techniques for navigation in a model-based virtual environment.

The final aspect related to Research Question 2 is the web-based implementation and navigation techniques. The system shows a web-based, light-weight representation of the product can be used for the navigation of information. While the Formula Student racing car is a far smaller and less complex product than an aircraft, the prototype shows the feasibility of the approach to fulfil the Airbus desire of a 'democratised' CAD model and making it available to all personnel with a web-browser rather than just a select few 'CAD engineers' with high-end computer hardware and software.

10.3.3 Research Question 3

Research Question 3 produced a set of visual information objects for the display of information in the model-based interface. Components-of-interest, region-of-interest, section-of-interest, and points-of-interest were designed and validated. A repair case relating to corrosion is one of the best examples of where these markers have a relevance to Airbus. Corrosion is not always bound by the dimensions of a component, nor does it always encompass that component in its entirety and neither does it belong to a single x, y, and z co-ordinate within the model-based co-ordinate system. As such, using the component itself, or a single point in space does not reflect the physical and functional nature of corrosion. A feature-of-interest visual information object does encompass these and as such is the logical technique for displaying information within engineering three-dimensional virtual environments, be that a model-based search system, CAD, or virtual or augmented reality systems. The implications of the third research question then span beyond information access remit of this thesis and impact emerging technologies that may someday be used within the organisation.

10.3.4 Research Question 4

The fourth and final research question investigated how a model-based interface compares to a more traditional text-based interface search engine. Key to success of the in-service team is finding similar past cases and relevant information early in the repair request process, therefore quick response times are important. The findings showed no significant difference in times for five of the nine tasks and two quicker and two slower for the model-based interface compared to the text-based interface. A number of participants showed a preference for having both systems side-by-side, if this were provided, then the In-Service team could benefit from the speedier completion times for the relevant tasks. This also assumes there will not be an improvement in completion times as users become familiar with the model-based interface develop search strategies, something that requires further investigation.

The user study also revealed that the likelihood of finding information is heavily dependent upon whether the participants were familiar with the product/project/in a leadership role or not. The findings also showed that for experienced users, the task's completion time was approximately similar to that of the maximum achieved by the text-based interface. Given Airbus products are in the Service stage of their life-cycle, the product is mature and as such, personnel are more likely to be familiar with the product hence more likely to find information. If this were not the case, the findings also show there would not be any significant negative impact on the total times to find information over a traditional text-based interface.

10.4 Thesis Reflections

This section reflects on the thesis and briefly covers the areas that would be altered if the thesis were started again.

10.4.1 Research Question 1

Research Question 1 determined that document classification against the product structure can be used to improve search of engineering documents. The findings show that this is possible, however, there is a lot more work that could be done in testing the various document classification techniques on text documents, as well as machine learning techniques and the various media (CAD files/drawings, etc.) that engineers use. The work presented in response to this research question would provide further robustness and a wider contribution to knowledge if this additional work had been performed.

It was during this study when the nature of their work resulted in Airbus Wing In-Service not being able to participate in this or future studies and as such, it was also at this point where the thesis began focusing on the Formula Student surrogates. Before being cancelled, the first indexing study was in progress at Airbus and was eliciting the similarity of In-Service reports directly from In-Service engineers. The aim being the identification of the attributes that the engineers considered when evaluating similar reports, such that the attributes could be leveraged for document indexing. Being able to identify this list of attributes would have been a significant contribution to knowledge as well as answering the research question at the same time. So while the first research question was answered, there was more that could have been contributed at the same time.

10.4.2 Research Question 2

Research Question 2 generated the range of navigation techniques. The findings presented in this section were qualitative and taken from the findings of the final study. The original plan was to perform a “Where’s Wally?” style study where participants would be asked to find a marker within the model space using the range of navigation techniques. The system would have logged the time taken and as such, allowed for a quantitative comparison of the techniques. The findings from Research Question 2 stands, however it is not possible to compare between navigation techniques and this would have made a wider and interesting contribution to knowledge. Incidentally, the “Where’s Wally?” study was not performed due to time restrictions and access to the

Formula Student teams.

10.4.3 Research Question 3

Research Question 3 generated the markers for identifying the location of information within the model-based virtual environment. As concluded in the relevant **Chapter 7: The Design of Visual Information Objects in Three-Dimensional Virtual Environments for Engineering Information Navigation**, there was some ambiguity between the directional and non-directional version of the point-of-interest marker. Following this up with further study would have provided a further completeness in understanding of the appropriateness of the point-of-interest marker usage. The limitation causes by the use of marker colour should also be eliminated to verify the findings presented. The time restrictions on the thesis mean that a second more detailed study was not possible.

10.4.4 Research Question 4

Research Question 4 determined the whether a model-based approach to information navigation improves information access and knowledge discovery. The findings state how some users would have preferred more time for familiarisation or stated that the text-based interface was more familiar. As stated in the relevant **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**, further findings could have been achieved if the system had been provided to the participants for a longer period of time (months) prior to the study being performed. This would have given participants the required time to fully familiarise themselves with the model-based interface.

Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results also described how the UWE participants were included alongside the participants from the other universities in spite of them seeing a slightly different model-based user interface. While at the time this were justified by the changes being minimal and the effect of averaging results across all participants removing most of the affect, this is still a limitation and further study on a consistent model-based interface would improve the results presented.

10.4.5 Other Areas of Work

The Formula Student racing car acted as a surrogate for a far larger and more complex artefact, i.e. an aircraft. During the formulation of the research questions an area considered that had a significant amount of time dedicated to it, was a technique for visualising these large and complex artefacts within the limited memory of the web-browser. A novel technique was designed that allowed to user to navigate the product structure by clicking on systems → subsystems → component. The first view of the entire aircraft showed high level systems in a low resolution rather than all components. To select an individual component, the user clicked on a high level system, all other systems disappeared and were replaced with the subsystems/components of the selected high level system. This continued all the way down the product structure. After the studies moved away from Airbus to the Formula Student surrogate, the size and complexity were no longer a problem and work moved away from this technique. The technique was however developed but it has not been robustly studied and the contribution presented. Adding this would have generated a greater contribution to knowledge.

10.5 Future Work

The framework that this thesis has generated covers a range of research fields and to bring all the findings together within the time limitations of the project, there were a number of areas that would benefit from further research. These range from improving the findings presented here to a host of new features and model-based concepts. This section now explores a number of these in turn with the aim of inspiring future research efforts in the area of model-based engineering information management.

10.5.1 Model-Based Information Navigation, Further Long-Term Studies

As touched upon in **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**, the Research Question 4 study showed how some participants felt they needed more time to become familiar with the system or that the text-based interface was overwhelmingly more familiar than the model-based interface. It is then understandable that longer-term studies are mentioned here as an area of future work in an attempt to mitigate this. The study would involve providing the system to engineers as part of their day-to-day operations for a period of time (six months for example) and then re-run this study. This should in effect eliminate any learning curves or bias towards the more familiar text-base interface by allowing participants to generate their own search strategies.

10.5.2 Investigating Model-Based Mixed Media Document Indexing

The findings from the first research question showed how document classification can be used to improve document indexing and that this allows for the indexing of a range of media. This thesis does not however explore the most appropriate techniques for delivering this within a real-world environment. This thesis then calls for future work to determine appropriate techniques for indexing the range of engineering document media.

In addition to this, work should also be performed in determining appropriate techniques to present those documents to the user. In Internet search, we are used to an interface that presents different media separately - when using Google for example, one selects the medium (images, video, map) before searching and being provided those results. Machine learning techniques of classification and case-based reasoning produce a measure of the relationship between the document and component, regardless of the media. There is then an opportunity to investigate when one could classify and rank documents against components independent of media.

10.5.3 Model-Based Information Navigation in the Project Design Phase

The use-cases explored in this thesis are based on an engineering product in the in-service phase of the product life-cycle. That is, once the product actually exist. This raises the question, how does such an approach work in the design phase if the product design is evolving and changing? Essentially, as a particular feature is removed from a design, how is that information accessed, or should it be accessed? There is then a question of whether a model-based approach to information navigation can be used during design and if so, how? Two possible approaches would be displaying the evolving CAD model and generic CAD model. The evolving CAD model approach gives the user control over the product evolution time-line allowing them to effectively move back in time to where that feature existed. The generic CAD model approach involves a model that encompasses a generic product model that changes to become the final model as it is designed. If, for example, one places a basic cube shape in place of a car engine, would engineers understand and interpret that as an engine ‘place holder’ such that it is usable for the access of information? There is a host of interesting and worthwhile science that could be done in this area.

10.5.4 Model-Based Information Visualisation, Project Management, and Communication

As presented in **Chapter 9: Does Model-Based Information Navigation Improve Information Access and Knowledge Discovery. Part 2: Study Results**, the heat map presented in this thesis as a means of displaying information and navigating the model-based environment is based on an information visualisation technique. The Language of Collaborative Manufacturing project (EPSRC Grant Reference EP/K014196/2), is investigating the use of the heat map in a project management ‘dashboard’. This thesis has shown the benefit of the heat map in seeing who worked on what and this technique has potential in answering other project level questions such as time to completion, activity, and problems. The basis of this thesis is that engineers thin visually and functionally and as such providing a visual and functional search interface should improve search. This premise stands for the other aspects of the product life-cycle, including the project itself.

There is then also the opportunity to investigate whether these additional aspects of the product life-cycle would benefit from a model-based approach. The final example of this discussed here is communication. The multinational nature of organisations such as Airbus results in teams being distributed across sites, countries and continents. There is a question of whether allowing engineers to communicate via the product structure could improve communication. The idea being that online-chat conversations take place in real-time and via the product structure and stored and made searchable. There is also scope here to investigate whether this allows for the capture of design rationale, although one must address whether a model-based approach is suitable during the design phase first.

An organisation’s data, information, and knowledge is widely considered to be one of its greatest assets. As such, the capture, storage and dissemination of this asset is the focus of research and organisational efforts. This thesis adds to this field through the development of a framework for model-based information navigation: a novel approach to finding information that places a three-dimensional representation of the product at the heart of navigating document collections.

Bibliography

[1] L. T. Blessing and A. Chakrabarti, *DRM, a design research methodology*. Springer Science & Business Media, 2009.

[2] B. J. Hicks, S. J. Culley, and C. A. McMahon, "A study of issues relating to information management across engineering smes," *International Journal of Information Management*, vol. 26, no. 4, pp. 267–289, 2006.

[3] P. Heisig, N. H. Caldwell, K. Grebici, and P. J. Clarkson, "Exploring knowledge and information needs in engineering from the past and for the future–results from a survey," *Design studies*, vol. 31, no. 5, pp. 499–532, 2010.

[4] T. H. Davenport and L. Prusak, *Working knowledge: How organizations manage what they know*. Harvard Business Press, 1998.

[5] J. Rowley, "The wisdom hierarchy: representations of the dikw hierarchy," *Journal of Information Science*, vol. 33, no. 2, pp. 163–180, 2007.

[6] Z. Li, V. Raskin, and K. Ramani, "Developing engineering ontology for information retrieval," *Journal of Computing and Information science in Engineering*, vol. 8, no. 1, p. 011003, 2008.

[7] C. McMahon, A. Lowe, S. Culley, M. Corderoy, R. Crossland, T. Shah, and D. Stewart, "Waypoint: an integrated search and retrieval system for engineering documents," *Journal of Computing and Information Science in Engineering*, vol. 4, no. 4, pp. 329–338, 2004.

[8] Y. Xie, *Application of context aware systems to support knowledge work in the aerospace*. PhD thesis, University of Bath, 2013.

[9] A. Stocker, M. Zoier, S. Softic, S. Paschke, H. Bischofter, and R. Kern, "Is enterprise search useful at all?: lessons learned from studying user behavior," in

Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business, p. 22, ACM, 2014.

[10] R. Mukherjee and J. Mao, "Enterprise search: Tough stuff," *Queue*, vol. 2, pp. 36–46, Apr. 2004.

[11] M. Alemanni, F. Destefanis, and E. Vezzetti, "Model-based definition design in the product lifecycle management scenario," *The International Journal of Advanced Manufacturing Technology*, vol. 52, pp. 1–14, Jan 2011.

[12] C. McMahon and D. Davies, "The use of annotation in design representation," *DESIGN METHODS FOR PRACTICE-Cover*, 2006.

[13] C. Li, C. McMahon, and L. Newnes, "Annotation in design processes: Classification of approaches," in *DS 58-8: Proceedings of ICED 09, the 17th International Conference on Engineering Design, Vol. 8, Design Information and Knowledge, Palo Alto, CA, USA, 24.-27.08. 2009*, 2009.

[14] N. Bojcetic, N. Pavkovic, D. Pavlic, *et al.*, "Extended cad model," in *ICED 05: 15th International Conference on Engineering Design: Engineering Design and the Global Economy*, p. 1731, Engineers Australia, 2005.

[15] C. M. Eastman, C. Eastman, P. Teicholz, and R. Sacks, *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors.* John Wiley & Sons, 2011.

[16] C. Marrin, "Webgl specification," *Khronos WebGL Working Group*, 2011.

[17] J. Dirksen, *Learning Three. js: the JavaScript 3D library for WebGL.* Packt Publishing Ltd, 2013.

[18] S. Quintana-Amate, P. Bermell-Garcia, and A. Tiwari, "Transforming expertise into knowledge-based engineering tools: A survey of knowledge sourcing in the context of engineering design," *Knowledge-Based Systems*, vol. 84, pp. 89–97, 2015.

[19] M. Hall, C. A. McMahon, P. Bermell-Garcia, A. Johansson, R. Ravindranath, *et al.*, "Capturing synchronous collaborative design activities: A state-of-the-art technology review," in *The International Design Conference-DESIGN 2018International Design Conference*, pp. 347–358, Design Society, 2018.

[20] E. Carey, *Visually Informed Support for Design Engineering Decisions.* PhD thesis, University of Bath, 2016.

[21] Y. Xie, S. Culley, and F. Weber, "Opportunities and challenges for context-aware systems in aerospace industry," *Journal of Enterprise Information Management*, vol. 24, no. 2, pp. 118–125, 2011.

[22] F. Weber, E. Dauphin, R. Fuschini, J. Haarmann, A. Katzung, and M. Wunram, "Expertise transfer: A case study about knowledge retention at airbus," in *Technology Management Conference (ICE), 2007 IEEE International*, pp. 1–10, IEEE, 2007.

[23] D. E. Jones, Y. Xie, C. McMahon, M. Dotter, N. Chanchevrier, and B. Hicks, "Improving enterprise wide search in large engineering multinationals: A linguistic comparison of the structures of internet-search and enterprise-search queries," in *Product Lifecycle Management in the Era of Internet of Things* (A. Bouras, B. Eynard, S. Foufou, and K.-D. Thoben, eds.), (Cham), pp. 216–226, Springer International Publishing, 2016.

[24] R. Redon, A. Larsson, R. Leblond, and B. Longueville, "Vivace context based search platform," in *Modeling and Using Context* (B. Kokinov, D. C. Richardson, T. R. Roth-Berghofer, and L. Vieu, eds.), (Berlin, Heidelberg), pp. 397–410, Springer Berlin Heidelberg, 2007.

[25] F. De Florio, *Airworthiness: An introduction to aircraft certification and operations*. Butterworth-Heinemann, 2016.

[26] R. L. Ackoff, "From data to wisdom," *Journal of applied systems analysis*, vol. 16, no. 1, pp. 3–9, 1989.

[27] Y. Zhao, L. C. Tang, M. J. Darlington, S. A. Austin, and S. J. Culley, "High value information in engineering organisations," *International Journal of Information Management*, vol. 28, no. 4, pp. 246–258, 2008.

[28] D. Leonard-Barton, "Wellsprings of knowledge: Building and sustaining the sources of innovation," 1995.

[29] S. L. Pan and H. Scarbrough, "Knowledge management in practice: An exploratory case study," *Technology Analysis & Strategic Management*, vol. 11, no. 3, pp. 359–374, 1999.

[30] M. Alavi and D. E. Leidner, "Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues," *MIS Quarterly*, vol. 25, no. 1, pp. 107–136, 2001.

[31] F. J. Forcadell and F. Guadamillas, "A case study on the implementation of a knowledge management strategy oriented to innovation," *Knowledge and Process Management*, vol. 9, pp. 162–171, 7 2002.

[32] Y. Yeh, S. Lai, and C. Ho, "Knowledge management enablers: a case study," *Industrial Management & Data Systems*, vol. 106, no. 6, pp. 793–810, 2006.

[33] Y.-C. Lin, "Construction 3d bim-based knowledge management system: a case study," *Journal of Civil Engineering and Management*, vol. 20, no. 2, pp. 186–200, 2014.

[34] M. S. Ackerman, V. Pipek, and V. Wulf, *Sharing expertise: Beyond knowledge management.* MIT press, 2003.

[35] I. Litvaj and D. Stancekova, "Decision - making, and their relation to the knowledge management, use of knowledge management in decision - making," *Procedia Economics and Finance*, vol. 23, pp. 467 – 472, 2015. 2nd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM.

[36] E. T. Penrose, *The Theory of the Growth of the Firm.* Oxford university press, 2009.

[37] M. W. Salisbury, "Putting theory into practice to build knowledge management systems," *Journal of Knowledge Management*, vol. 7, no. 2, pp. 128–141, 2003.

[38] C. Wagner and N. Bolloju, "Supporting knowledge management in organizations with conversational technologies: Discussion forums, weblogs, and wikis," *Journal of Database Management*, vol. 16, no. 2, p. I, 2005.

[39] B. Hicks, S. Culley, R. Allen, and G. Mullineux, "A framework for the requirements of capturing, storing and reusing information and knowledge in engineering design," *International Journal of Information Management*, vol. 22, no. 4, pp. 263 – 280, 2002.

[40] A.-H. Tan *et al.*, "Text mining: The state of the art and the challenges," in *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, vol. 8, pp. 65–70, sn, 1999.

[41] Y. Xie, S. Culley, and F. Weber, "Applying context to organize unstructured information in aerospace industry," in *DS 68-6: Proceedings of the 18th International*

Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 6: Design Information and Knowledge, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011, 2011.

[42] R. Bracewell, K. Wallace, M. Moss, and D. Knott, "Capturing design rationale," *Computer-Aided Design*, vol. 41, no. 3, pp. 173–186, 2009.

[43] L. Lan, Y. Liu, and W. F. Lu, "Learning from the past: Uncovering design process models using an enriched process mining," *Journal of Mechanical Design*, vol. 140, no. 4, p. 041403, 2018.

[44] T. J. Howard, S. Culley, and E. Dekoninck, "Information as an input into the creative process," in *DS 36: Proceedings DESIGN 2006, the 9th International Design Conference, Dubrovnik, Croatia*, 2006.

[45] M. Giess and S. Culley, "Investigating manufacturing data for use within design," in *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm*, 2003.

[46] S. Liu, C. A. McMahon, and S. J. Culley, "A review of structured document retrieval (sdr) technology to improve information access performance in engineering document management," *Computers in Industry*, vol. 59, no. 1, pp. 3–16, 2008.

[47] S. Liu, C. A. McMahon, M. J. Darlington, S. J. Culley, and P. J. Wild, "A computational framework for retrieval of document fragments based on decomposition schemes in engineering information management," *Advanced Engineering Informatics*, vol. 20, no. 4, pp. 401–413, 2006.

[48] I. Becerra-Fernandez and R. Sabherwal, *Knowledge management: Systems and processes.* Routledge, 2014.

[49] P. Tyndale, "A taxonomy of knowledge management software tools: origins and applications," *Evaluation and Program Planning*, vol. 25, no. 2, pp. 183 – 190, 2002.

[50] F. Ameri and D. Dutta, "Product lifecycle management: closing the knowledge loops," *Computer-Aided Design and Applications*, vol. 2, no. 5, pp. 577–590, 2005.

[51] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*, vol. 39. Cambridge University Press, 2008.

[52] J. Duffy, "The km technology infrastructure," *Information Management*, vol. 34, no. 2, p. 62, 2000.

[53] J. Duffy, "Knowledge management: What every information professional should know," *Information Management*, vol. 34, no. 3, p. 10, 2000.

[54] J. Duffy, "The tools and technologies needed for knowledge management," *Information Management*, vol. 35, no. 1, p. 64, 2001.

[55] J. Duffy, "Something funny is happening on the way to knowledge management...," *Information Management*, vol. 34, no. 4, p. 64, 2000.

[56] I. Muthalagu, "Importance of knowledge management in engineering industries," 2017.

[57] S. A. A. Saifi, S. Dillon, and R. McQueen, "The relationship between face to face social networks and knowledge sharing: an exploratory study of manufacturing firms," *Journal of Knowledge Management*, vol. 20, no. 2, pp. 308–326, 2016.

[58] Y. Li, Z. Liu, and H. Zhu, "Enterprise search in the big data era: Recent developments and open challenges," *Proc. VLDB Endow.*, vol. 7, pp. 1717–1718, Aug. 2014.

[59] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web.," tech. rep., Stanford InfoLab, 1999.

[60] A. Gilchrist, "Thesauri, taxonomies and ontologies – an etymological note," *Journal of Documentation*, vol. 59, no. 1, pp. 7–18, 2003.

[61] K. H. Han and J. W. Park, "Process-centered knowledge model and enterprise ontology for the development of knowledge management system," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7441 – 7447, 2009.

[62] Y. Xie, *Application of context aware systems to support knowledge work in the aerospace.* PhD thesis, University of Bath, 2013.

[63] L. Li, S. Gao, Y. Liu, and X. Qin, "Enhanced sparql-based design rationale retrieval," *AI EDAM*, vol. 30, no. 4, pp. 406–423, 2016.

[64] D. Campbell, S. Culley, C. McMahon, *et al.*, "'push-based'strategies for improving the efficiency of information management in design," in *ICED 05: 15th International Conference on Engineering Design: Engineering Design and the Global Economy*, p. 428, Engineers Australia, 2005.

[65] Y. Baba and K. Nobeoka, "Towards knowledge-based product development: the 3-d cad model of knowledge creation," *Research Policy*, vol. 26, no. 6, pp. 643 – 659, 1998.

[66] V. Quintana, L. Rivest, R. Pellerin, F. Venne, and F. Kheddouci, "Will model-based definition replace engineering drawings throughout the product lifecycle? a global perspective from aerospace industry," *Computers in Industry*, vol. 61, no. 5, pp. 497 – 508, 2010.

[67] A. K. M. Team, "Airbus knowledge management service portfolio," 2014.

[68] P. T. AG, "Business search v3 search log analytics," 2014.

[69] D. J. Power, R. Sharda, and F. Burstein, *Decision support systems.* Wiley Online Library, 2015.

[70] BS-1192-4:2014, "Collaborative production of information part 4: Fulfilling employer's information exchange requirements using cobie – code of practice," tech. rep., British Standards Institution, September 2014.

[71] M. Garetti, S. Terzi, N. Bertacci, and M. Brianza, "Organisational change and knowledge management in plm implementation," *International Journal of Product Lifecycle Management*, vol. 1, no. 1, pp. 43–51, 2005.

[72] J. Stark, "Product lifecycle management," in *Product Lifecycle Management (Volume 2)*, pp. 1–35, Springer, 2016.

[73] R. J. Rost, B. Licea-Kane, D. Ginsburg, J. Kessenich, B. Lichtenbelt, H. Malan, and M. Weiblen, *OpenGL shading language.* Pearson Education, 2009.

[74] D. E. Jones, N. Chanchevrier, C. McMahon, and B. Hicks, "A strategy for artefact-based information navigation in large engineering organisations," in *DS 80-10 Proceedings of the 20th International Conference on Engineering Design (ICED 15) Vol 10: Design Information and Knowledge Management Milan, Italy, 27-30.07. 15*, 2015.

[75] T. Catarci and S. Kimani, "Human-computer interaction view on information retrieval evaluation," in *Information Retrieval Meets Information Visualization*, pp. 48–75, Springer, 2013.

[76] P.-N. Tan, M. Steinbach, V. Kumar, *et al.*, *Introduction to data mining*, vol. 1. Pearson Addison Wesley Boston, 2006.

[77] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 249–256, ACM, 1990.

[78] C. Wharton, J. Rieman, C. Lewis, and P. Polson, "The cognitive walkthrough method: A practitioner's guide," in *Usability inspection methods*, pp. 105–140, John Wiley & Sons, Inc., 1994.

[79] J. R. Lewis, "Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use," *International Journal of Human-Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.

[80] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2016.

[81] S. Mizzaro, "Relevance: The whole history," *Journal of the Association for Information Science and Technology*, vol. 48, no. 9, pp. 810–832, 1997.

[82] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval.* Cambridge University Press, 2008.

[83] B. Huang and Z. Xia, "Allocating inverted index into flash memory for search engines," in *Proceedings of the 20th international conference companion on World wide web*, pp. 61–62, ACM, 2011.

[84] M. Catena, C. Macdonald, and I. Ounis, "On inverted index compression for search engine efficiency," in *Advances in Information Retrieval* (M. de Rijke, T. Kenter, A. P. de Vries, C. Zhai, F. de Jong, K. Radinsky, and K. Hofmann, eds.), (Cham), pp. 359–371, Springer International Publishing, 2014.

[85] B. B. Cambazoglu, E. Kayaaslan, S. Jonassen, and C. Aykanat, "A term-based inverted index partitioning model for efficient distributed query processing," *ACM Trans. Web*, vol. 7, pp. 15:1–15:23, Sept. 2013.

[86] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.", 2009.

[87] S. Bird and E. Loper, "Nltk: the natural language toolkit," in *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, p. 31, Association for Computational Linguistics, 2004.

[88] J. Teevan, S. T. Dumais, and E. Horvitz, "Personalizing search via automated analysis of interests and activities," *SIGIR Forum*, vol. 51, pp. 10–17, Feb. 2018.

[89] E. Agichtein, E. Brill, and S. Dumais, "Improving web search ranking by incorporating user behavior information," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, (New York, NY, USA), pp. 19–26, ACM, 2006.

[90] M. Giess, P. J. Wild, and C. A. McMahon, "The generation of faceted classification schemes for use in the organisation of engineering design documents," *International journal of information management*, vol. 28, no. 5, pp. 379–390, 2008.

[91] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[92] P.-N. Tan *et al.*, *Introduction to data mining*. Pearson Education India, 2007.

[93] R. Baeza-Yates, B. Ribeiro-Neto, *et al.*, *Modern information retrieval*, vol. 463. ACM press New York, 1999.

[94] A. Dix, "Human-computer interaction," in *Encyclopedia of database systems*, pp. 1327–1331, Springer, 2009.

[95] J. Johnson, T. L. Roberts, W. Verplank, D. C. Smith, C. H. Irby, M. Beard, and K. Mackey, "The xerox star: A retrospective," *Computer*, vol. 22, no. 9, pp. 11–26, 1989.

[96] J. Tidwell, *Designing interfaces: Patterns for effective interaction design.* " O'Reilly Media, Inc.", 2010.

[97] R. A. Bolt, *"Put-that-there": Voice and gesture at the graphics interface*, vol. 14. ACM, 1980.

[98] R. S. Cooper, J. F. McElroy, W. Rolandi, D. Sanders, R. M. Ulmer, and E. Peebles, "Personal virtual assistant," June 29 2004. US Patent 6,757,362.

[99] T. R. Gruber, "Siri, a virtual personal assistant—bringing intelligence to the interface," 2009.

[100] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev, *3D user interfaces: Theory and practice.* Addison-Wesley Professional, 2017.

[101] B. Gerkey, R. T. Vaughan, and A. Howard, "The player/stage project: Tools for multi-robot and distributed sensor systems," in *Proceedings of the 11th international conference on advanced robotics*, vol. 1, pp. 317–323, 2003.

[102] J. B. Dabney and T. L. Harman, *Mastering simulink*. Pearson, 2004.

[103] P. Pirolli and S. Card, "Information foraging.," *Psychological review*, vol. 106, no. 4, p. 643, 1999.

[104] J. Blake and H. B. Gurocak, "Haptic glove with mr brakes for virtual reality," *IEEE/ASME Transactions On Mechatronics*, vol. 14, no. 5, pp. 606–615, 2009.

[105] T. Carter, S. A. Seah, B. Long, B. Drinkwater, and S. Subramanian, "Ultrahaptics: multi-point mid-air haptic feedback for touch surfaces," in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pp. 505–514, ACM, 2013.

[106] A. M. Okamura, J. T. Dennerlein, and R. D. Howe, "Vibration feedback models for virtual environments," in *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, vol. 1, pp. 674–679, IEEE, 1998.

[107] HTC Corporation, "Vive user guide." `http://dl4.htc.com/web_materials/Manual/Vive/Vive_User_Guide.pdf`, 2017. Accessed:2017-06-29.

[108] J. Jankowski and M. Hachet, "A survey of interaction techniques for interactive 3d environments," in *Eurographics 2013-STAR*, 2013.

[109] J. D. Mackinlay, S. K. Card, and G. G. Robertson, "Rapid controlled movement through a virtual 3d workspace," in *ACM SIGGRAPH computer graphics*, vol. 24, pp. 171–176, ACM, 1990.

[110] A. Khan, I. Mordatch, G. Fitzmaurice, J. Matejka, and G. Kurtenbach, "Viewcube: a 3d orientation indicator and controller," in *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, pp. 17–25, ACM, 2008.

[111] M. Brereton and B. McGarry, "An observational study of how objects support engineering design thinking and communication: implications for the design of tangible media," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 217–224, ACM, 2000.

[112] M. R. Mine, "Virtual environment interaction techniques," *UNC Chapel Hill CS Dept*, 1995.

[113] G. Fitzmaurice, A. Khan, R. Pieké, B. Buxton, and G. Kurtenbach, "Tracking menus," in *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pp. 71–79, ACM, 2003.

[114] G. Fitzmaurice, J. Matejka, I. Mordatch, A. Khan, and G. Kurtenbach, "Safe 3d navigation," in *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, pp. 7–15, ACM, 2008.

[115] G. Svennerberg, *Beginning Google Maps API 3.* Apress, 2010.

[116] R. Mohammad and E. Kroll, "Automatic generation of exploded view by graph transformation," in *Artificial Intelligence for Applications, 1993. Proceedings., Ninth Conference on*, pp. 368–374, IEEE, 1993.

[117] W. Li, M. Agrawala, B. Curless, and D. Salesin, "Automated generation of interactive 3d exploded view diagrams," in *ACM Transactions on Graphics (TOG)*, vol. 27, p. 101, ACM, 2008.

[118] W. Li, L. Ritter, M. Agrawala, B. Curless, and D. Salesin, "Interactive cutaway illustrations of complex 3d models," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, p. 31, 2007.

[119] S. Knödel, M. Hachet, and P. Guitton, "Interactive generation and modification of cutaway illustrations for polygonal models," in *Smart Graphics*, pp. 140–151, Springer, 2009.

[120] S. Few, "Common pitfalls in dashboard design," *Perceptual Edge*, 2006.

[121] B. Shneiderman, *Designing the user interface: strategies for effective human-computer interaction*, vol. 3. Addison-Wesley Reading, MA, 1992.

[122] W. O. Galitz, *The essential guide to user interface design: an introduction to GUI design principles and techniques.* John Wiley & Sons, 2007.

[123] N. Iliinsky and J. Steele, *Designing Data Visualizations: Representing Informational Relationships.* " O'Reilly Media, Inc.", 2011.

[124] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[125] M. Maguire and B. Delahunt, "Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars.," *AISHE-J: The All Ireland Journal of Teaching and Learning in Higher Education*, vol. 9, no. 3, 2017.

[126] S. E. W. A. G. Department for Transport, Department for Regional Development (Northern Ireland), *Traffic Signs Manual (Eighth impression 2009).* Her Majesty's Stationery Office, 2009.

[127] S. R. Ellis, "Nature and origins of virtual environments: A bibliographical essay," *Computing Systems in Engineering*, vol. 2, no. 4, pp. 321–347, 1991.

[128] S. R. Ellis, "What are virtual environments?," *IEEE Computer Graphics and Applications*, vol. 14, no. 1, pp. 17–22, 1994.

[129] S. Kim and D. Weissmann, "Middleware-based integration of multiple cad and pdm systems into virtual reality environment," *Computer-Aided Design and Applications*, vol. 3, no. 5, pp. 547–556, 2006.

[130] S. Jayaram, U. Jayaram, Y. Wang, H. Tirumali, K. Lyons, and P. Hart, "Vade: A virtual assembly design environment," *IEEE Computer Graphics and Applications*, vol. 19, no. 6, pp. 44–50, 1999.

[131] J. Maxfield, T. Fernando, and P. Dew, "A distributed virtual environment for concurrent engineering," in *Virtual Reality Annual International Symposium, 1995. Proceedings.*, pp. 162–170, IEEE, 1995.

[132] F. Doil, W. Schreiber, T. Alt, and C. Patron, "Augmented reality for manufacturing planning," in *Proceedings of the workshop on Virtual environments 2003*, pp. 71–76, ACM, 2003.

[133] F. Liarokapis, N. Mourkoussis, M. White, J. Darcy, M. Sifniotis, P. Petridis, A. Basu, and P. F. Lister, "Web3d and augmented reality to support engineering education," *World Transactions on Engineering and Technology Education*, vol. 3, no. 1, pp. 11–14, 2004.

[134] A. Webster, S. Feiner, B. MacIntyre, W. Massie, and T. Krueger, "Augmented reality in architectural construction, inspection and renovation," in *Proc. ASCE Third Congress on Computing in Civil Engineering*, pp. 913–919, 1996.

[135] "BS 8888:2004," standard, British Standards Institution, 2005.

[136] M. Vaaraniemi, M. Freidank, and R. Westermann, "Enhancing the visibility of labels in 3d navigation maps," in *Progress and New Trends in 3D Geoinformation Sciences*, pp. 23–40, Springer, 2013.

[137] M. Tatzgern, D. Kalkofen, R. Grasset, and D. Schmalstieg, “Hedgehog labeling: View management techniques for external labels in 3d space,” in *Virtual Reality (VR), 2014 iEEE*, pp. 27–32, IEEE, 2014.

[138] B. Bell, S. Feiner, and T. Höllerer, “View management for virtual and augmented reality,” in *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pp. 101–110, ACM, 2001.

[139] W. K. Horton, *The icon book: Visual symbols for computer systems and documentation*. John Wiley & Sons, Inc., 1994.

[140] B. J. Hicks, S. J. Culley, R. Allen, and G. Mullineux, “A framework for the requirements of capturing, storing and reusing information and knowledge in engineering design,” *International journal of information management*, vol. 22, no. 4, pp. 263–280, 2002.

[141] B. Chadha, “Engineering information management to support enterprise business processes,” 2001.

[142] R. Allen, B. Hicks, and S. Culley, “Integrating electronic information for the design of mechanical systems: the designer’s perspective,” in *World Multiconference on Systemics, Cybernetics and Informatics*, pp. 266–271, University of Bath, 2000.

[143] H. McAlpine, B. J. Hicks, G. Huet, and S. J. Culley, “An investigation into the use and content of the engineer’s logbook,” *Design Studies*, vol. 27, no. 4, pp. 481–504, 2006.

[144] Y. Pnueli and E. Zussman, “Evaluating the end-of-life value of a product andimproving it by redesign,” *International Journal of Production Research*, vol. 35, no. 4, pp. 921–942, 1997.

[145] D. Hawking, “Challenges in enterprise search,” in *Proceedings of the 15th Australasian database conference-Volume 27*, pp. 15–24, Australian Computer Society, Inc., 2004.

[146] D. E. Jones, Y. Xie, C. McMahon, M. Dotter, N. Chanchevrier, and B. Hicks, “Improving enterprise wide search in large engineering multinationals: A linguistic comparison of the structures of internet-search and enterprise-search queries,” in *IFIP International Conference on Product Lifecycle Management*, pp. 216–226, Springer, 2015.

[147] J. Stark, “Product lifecycle management,” in *Product Lifecycle Management*, pp. 1–29, Springer, 2015.

[148] B. Shneiderman, *Designing the user interface: strategies for effective human-computer interaction.* Pearson Education India, 2010.

[149] S. L. Smith and J. N. Mosier, “Design guidelines for user-system interface software,” tech. rep., DTIC Document, 1984.

[150] J. L. Gabbard, I. J Edward Swan, and D. Hix, “The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality,” *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 1, pp. 16–32, 2006.

[151] M. Dewar, *Getting Started with D3: Creating Data-Driven Documents.* ” O’Reilly Media, Inc.”, 2012.

[152] K. Hartmann, T. Götzelmann, K. Ali, and T. Strothotte, “Metrics for functional and aesthetic label layouts,” in *Smart Graphics*, pp. 115–126, Springer, 2005.

[153] M. Ward, “A survey of engineers in their information world,” *Journal of Librarianship and Information Science*, vol. 33, no. 4, pp. 168–176, 2001.

[154] M. A. Robinson, “How design engineers spend their time: Job content and task satisfaction,” *Design Studies*, vol. 33, no. 4, pp. 391–425, 2012.

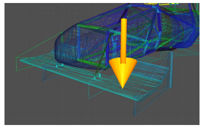
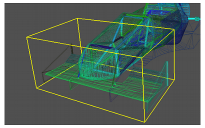
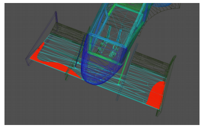
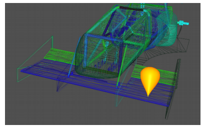
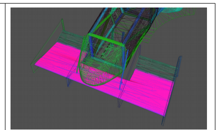
Appendix A

Appendices

A.1 Visual Information Object Quationnaire

Looking at the screen shots below, rank the markers from most (1) to least (5) appropriate base on whether you would use the marker for the association of the information source with the model.

Name:

					
Competition guidelines regarding a part/system/sub-system					
Details of the structural testing for a part/system/sub-system					
Damage report for a hole in a part/system/sub-system					
Technical details for a part/system/sub-system alignment adjustment					
Aerodynamics report for a part/system/sub-system					
Damage report showing locations effected by corrosion on a particular part/system/sub-system					
Livery and paint specifications for specific part/system/sub-system					

A.2 CSUQ Quationnaire (modified)

1 Participant Details

1. Please enter the Study ID displayed by the system.

2. How old are you?

3. What is your education level? (Undergraduate, BSc, MSc, etc.)

4. What is your degree program?

5. Do you have experience using 3D CAD and if so, how many years?

6. What is the extent of your Automotive Engineering Experience ? (years)

7. How long have you been involved with the Formula Student? (years)

8. What is your role within the Formula Student Team? and are you in a leadership role?

9. What are of the car are you most familiar with?

10. What type(s) of work are you involved with? (design, testing, analysis, leadership, etc)

1

2 System Usability Questionnaire

This questionnaire gives you an opportunity to tell us your reactions to the system you used. Your responses will help us understand what aspects of the system you are particularly concerned about and the aspects that satisfy you. To as great a degree as possible, think about all the tasks that you have done with the system while you answer these questions. Please read each statement and indicate how strongly you agree or disagree with the statement by circling a number on the scale. If a statement does not apply to you, circle N/A. Please write comments to elaborate on your answers. Thank you!

1. Overall, I am satisfied with how easy it is to use this system.

$StronglyAgree \leftarrow - \rightarrow StronglyDisagree$

1 2 3 4 5 6 7 N/A

Comments:

2. It was simple to use this system.

$StronglyAgree \leftarrow - \rightarrow StronglyDisagree$

1 2 3 4 5 6 7 N/A

Comments:

3. I could effectively complete the tasks and scenarios using this system.

$StronglyAgree \leftarrow - \rightarrow StronglyDisagree$

1 2 3 4 5 6 7 N/A

Comments:

4. I was able to complete the tasks and scenarios quickly using this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

5. I was able to efficiently complete the tasks and scenarios using this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

6. I felt comfortable using this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

7. It was easy to learn to use this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

8. I believe I could become productive quickly using this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

9. The interface of this system was pleasant.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

3

10. I liked using the interface of this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

11. This system has all the functions and capabilities I expect it to have.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

12. Overall, I am satisfied with this system.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

13. Please use this space for any further comments that you would like to make.

4

3 System Preference Questionnaire

This section gives you the opportunity to tell us your preference between the two types of search for engineering documents. If a statement does not apply to you, circle N/A. Please write comments to elaborate on your answers. Thank you!

1. I prefer the 3D model search over the text based search engine.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

2. For use with engineering information, I would prefer a 3D model search interface over text based search engines such as Google/SharePoint search.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

3. I find the 3D interface more intuitive to use than a text based search engine.

StronglyAgree ← − → *StronglyDisagree*

1 2 3 4 5 6 7 N/A

Comments:

4. Compared to a text based search engine, the 3D model interface improves my understanding of the presented search results.

StronglyAgree ← − → *StronglyDisagree*

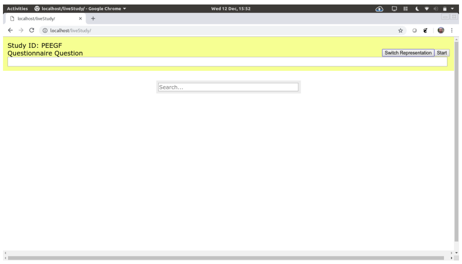
1 2 3 4 5 6 7 N/A

Comments:

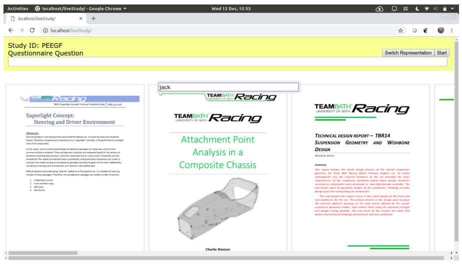
5. Please use this space for any further comments that you would like to make.

A.3 User Journeys

A.3.1 Text-Based Search Engine

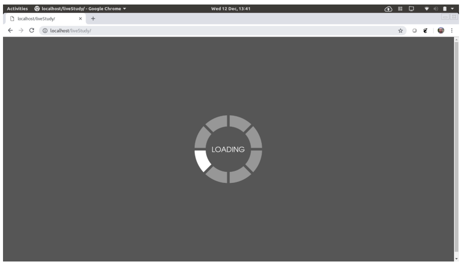


(a) Initial view of the text-based search engine

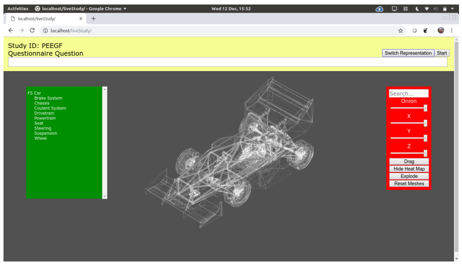


(b) The results panel from the text-based search engine. Clicking on a report front page opens up the report in another browser tab.

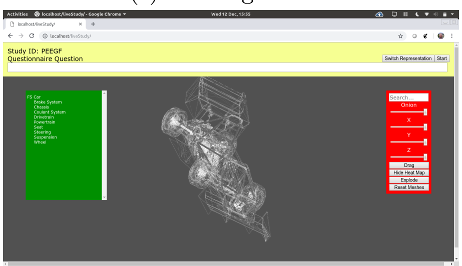
A.3.2 Model-Based System



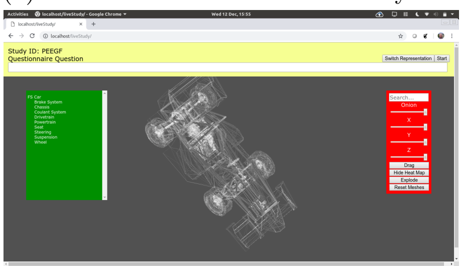
(c) Loading Screen



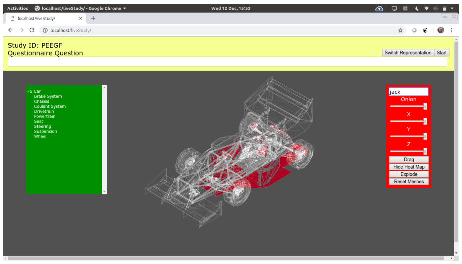
(d) Initial view of the model-based system

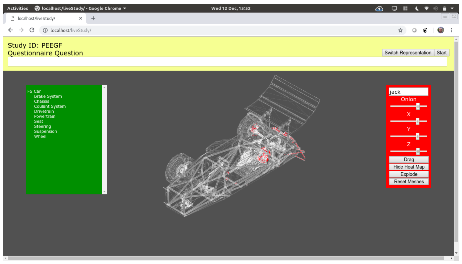


(e) Navigating the virtual environment by 'grabbing' and 'rotating' using the mouse.



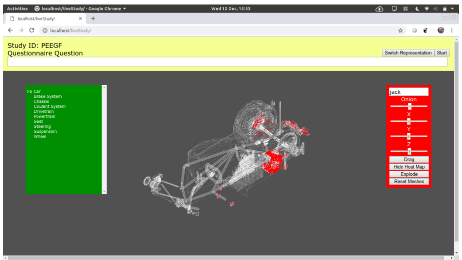
(f) Further navigation of the virtual environment.

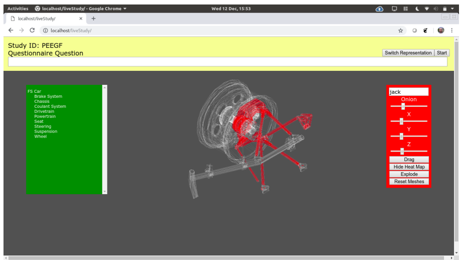




(g) Using the heat-map search facility within the red panel on the right. Searching for a term highlights the area of the car with documents corresponding to the query. Darker shades represent a stronger relationship.

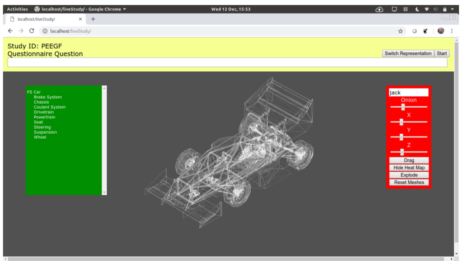
(h) Onion peeling through the car using the controls in the red panel on the right. Components are removed allowing access to the inner components.

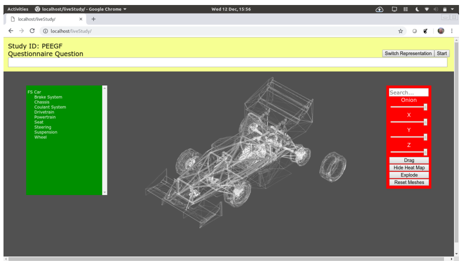




(i) Further onion peeling.

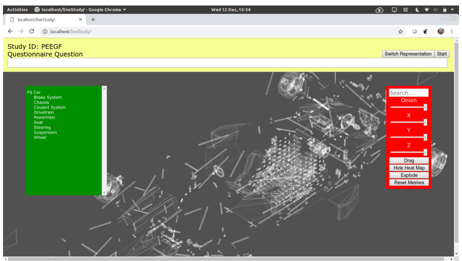
(j) Further onion peeling.



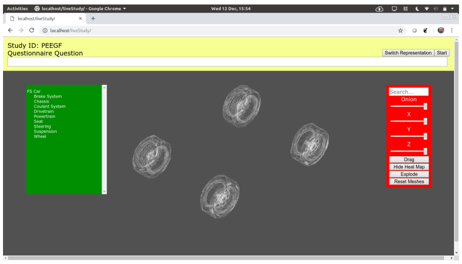


(k) Re-set mesh in the red-panel on the right re-sets the virtual environment.

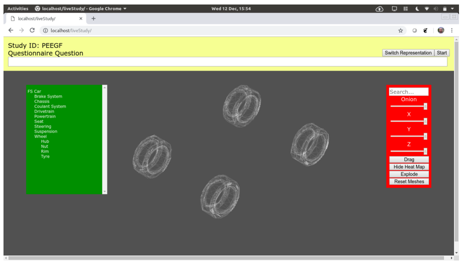
(l) Using the manual manipulation of components, enabled in the red panel on the right. Using the mouse, the user grabs and moves a back wheel.



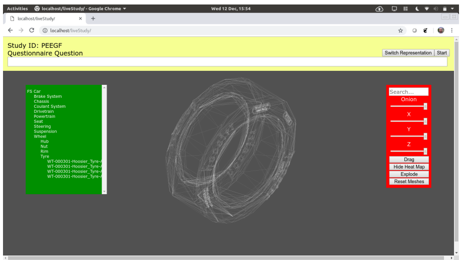
(m) The exploded view, enabled by clicking the 'Exploded View' button in the red panel on the right.



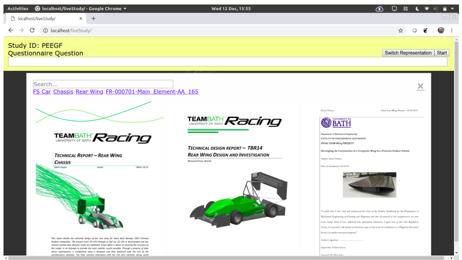
(n) Selecting systems/subsystems using the green panel on the left. Here the wheels are selected.



(o) Further selection of systems/subsystems using the green panel on the left. Here the tyres are selected.



(p) Further selection of a system/subsystem using the green panel on the left. Here an individual tyre is selected.

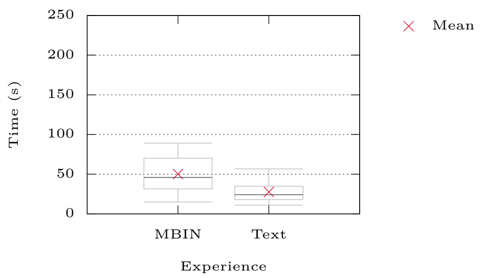


(q) The results panel from the model-based system showing results for a particular component. This is activated by double clicking on a component. Clicking on a report front page opens up the report in another browser tab.

A.4 Participant Formula Student Experience

A.4.1 Task 1: Analyse Requirements

Less than 2.5 years

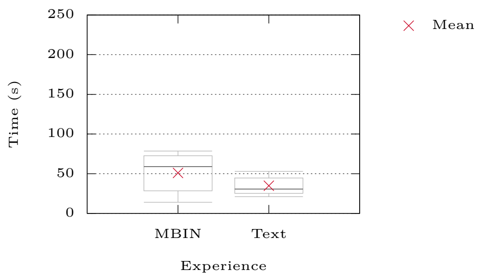


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	16.00	16.00
Mean	50.02	27.63
Standard Deviation	22.41	12.98
Skew	0.07	0.74
Skew Standard Error	0.56	0.56
Kurtosis	-1.34	-0.49
Kurtosis Standard Error	1.09	1.09
Shapiro-Wilk	0.94	0.92
- p-value	0.30	0.18
Significance		
T-Test		3.05
- p-value		0.01

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years

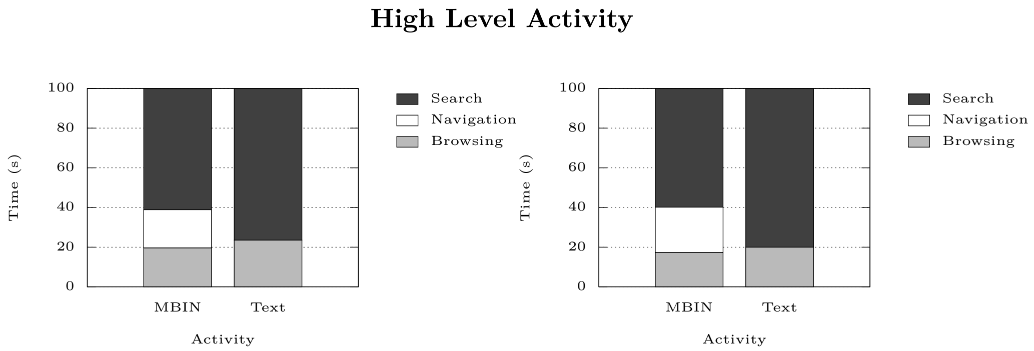


(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	6.00	6.00
Mean	51.01	34.73
Standard Deviation	25.26	11.84
*Sample too small for statistical comparison		

(d) Statistics for the Total Time for question completion for both interfaces

Figure A.2: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

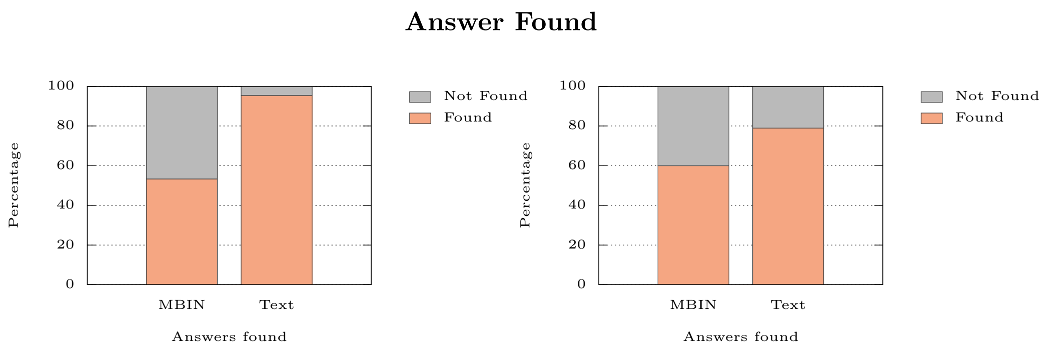


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.3: High level activity for below (a) and above (b) 2.5 years experience

(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.4: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

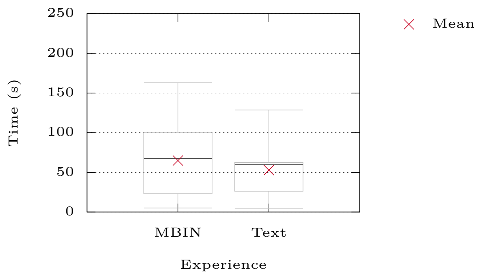


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.5: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.2 Task 2: Evaluate Last Year Sub-System Design; Approve Design

Less than 2.5 years

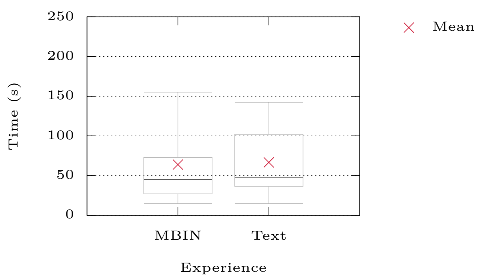


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	17.00	17.00
Mean	64.96	52.72
Standard Deviation	44.57	32.62
Skew	0.40	0.61
Skew Standard Error	0.55	0.55
Kurtosis	-0.71	-0.32
Kurtosis Standard Error	1.06	1.06
Shapiro-Wilk	0.95	0.94
- p-value	0.42	0.35
Significance		
T-Test		1.01
- p-value		0.33

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years



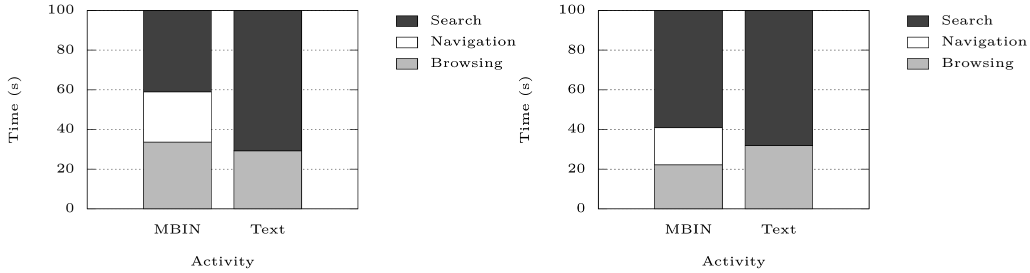
(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	9.00	9.00
Mean	63.89	66.63
Standard Deviation	47.61	44.04
Skew	0.86	0.48
Skew Standard Error	0.72	0.72
Kurtosis	-0.71	-1.20
Kurtosis Standard Error	1.40	1.40
Shapiro-Wilk	0.85	0.91
- p-value	0.07	0.31
Significance		
T-Test		-0.15
- p-value		0.88

(d) Statistics for the Total Time for question completion for both interfaces

Figure A.6: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

High Level Activity



(a) Less than 2.5 years (b) Greater than 2.5 years

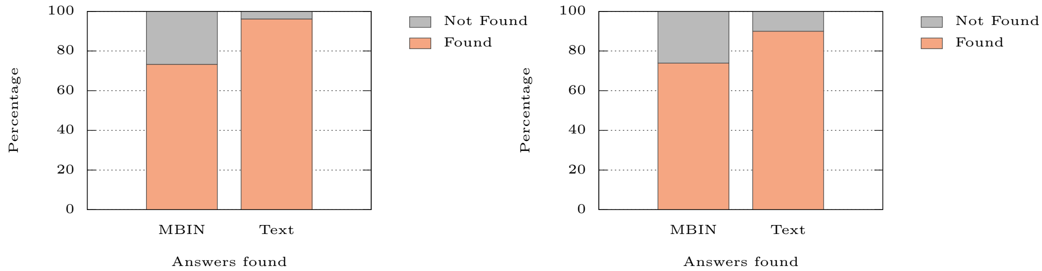
Figure A.7: High level activity for below (a) and above (b) 2.5 years experience

Reports Opened

(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.8: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

Answer Found

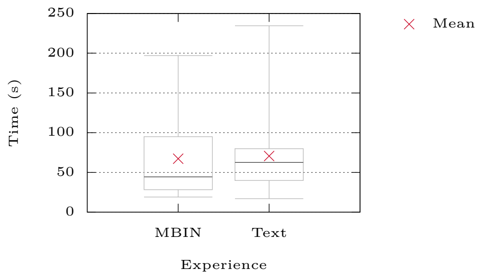


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.9: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.3 Task 3: Identify the Area of the Car Containing the Most Information

Less than 2.5 years



(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	16.00	16.00
Mean	67.22	70.54
Standard Deviation	49.83	51.67
Skew	1.20	1.88
Skew Standard Error	0.56	0.56
Kurtosis	0.59	3.66
Kurtosis Standard Error	1.09	1.09
Shapiro-Wilk	0.85	0.80
- p-value	0.01	0.00
Significance		
Wilcoxon		59.0
- p-value		0.64

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years

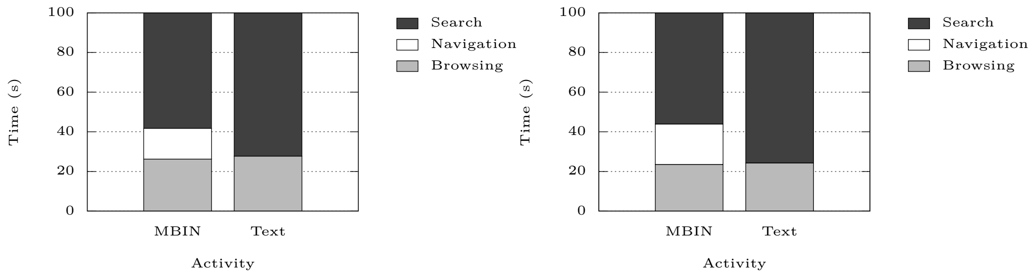
(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	12.00	12.00
Mean	40.17	47.97
Standard Deviation	26.36	24.99
Skew	0.84	0.98
Skew Standard Error	0.64	0.64
Kurtosis	-0.82	-0.13
Kurtosis Standard Error	1.23	1.23
Shapiro-Wilk	0.82	0.86
- p-value	0.02	0.05
Significance		
Wilcoxon		32.0
- p-value		0.58

(b) Statistics for the Total Time for question completion for both interfaces

Figure A.11: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

High Level Activity



(a) Less than 2.5 years (b) Greater than 2.5 years

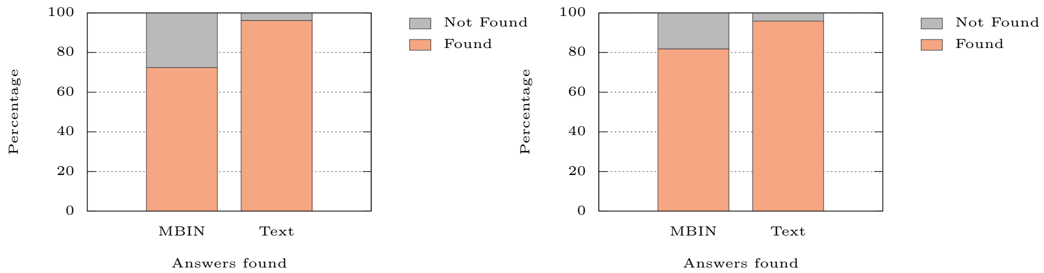
Figure A.12: High level activity for below (a) and above (b) 2.5 years experience

Reports Opened

(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.13: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

Answer Found

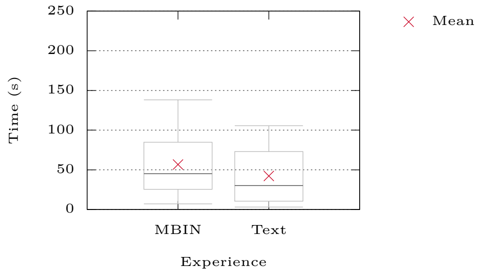


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.14: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.4 Task 4: Order Item

Less than 2.5 years

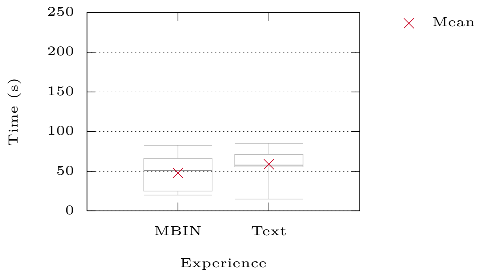


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	14.00	14.00
Mean	56.65	42.27
Standard Deviation	41.80	35.43
Skew	0.85	0.53
Skew Standard Error	0.60	0.60
Kurtosis	-0.66	-1.31
Kurtosis Standard Error	1.15	1.15
Shapiro-Wilk	0.86	0.85
- p-value	0.03	0.03
Significance		
Wilcoxon		35.0
- p-value		0.27

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years



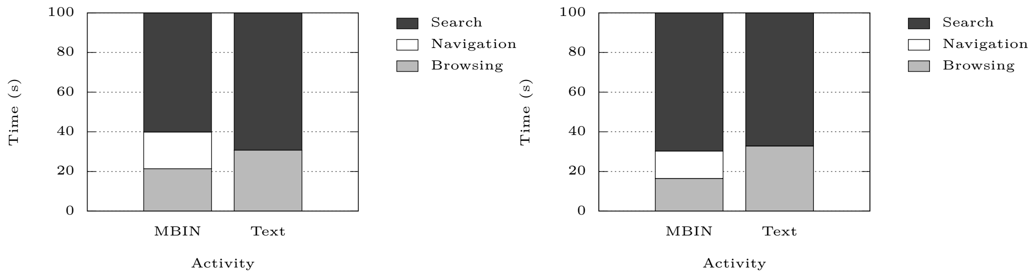
(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	7.00	7.00
Mean	47.94	58.89
Standard Deviation	24.30	20.83
Skew	0.30	-0.81
Skew Standard Error	0.79	0.79
Kurtosis	-1.44	0.11
Kurtosis Standard Error	1.59	1.59
Shapiro-Wilk	0.87	0.89
- p-value	0.18	0.27
Significance		
T-Test		-0.88
- p-value		0.41

(d) Statistics for the Total Time for question completion for both interfaces

Figure A.15: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

High Level Activity



(a) Less than 2.5 years (b) Greater than 2.5 years

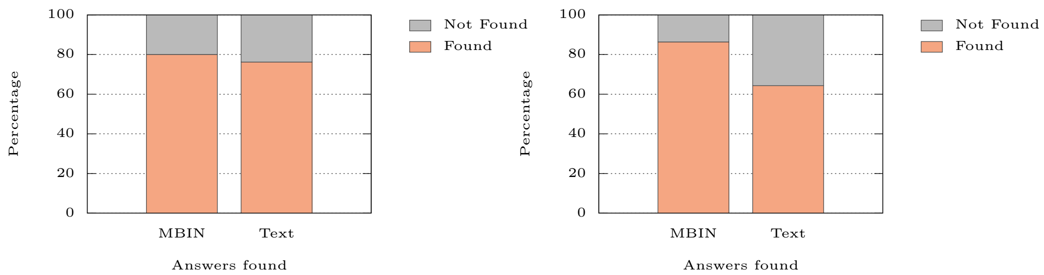
Figure A.16: High level activity for below (a) and above (b) 2.5 years experience

Reports Opened

(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.17: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

Answer Found

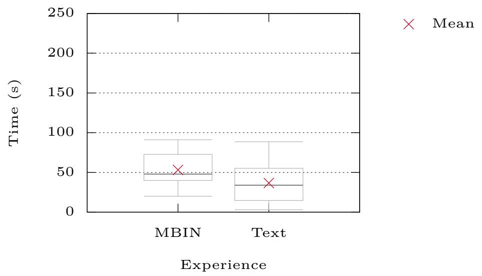


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.18: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.5 Task 5: Provide Information; Approve for Use of New Material

Less than 2.5 years

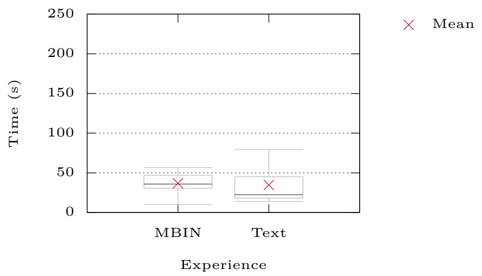


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	13.00	13.00
Mean	53.15	36.70
Standard Deviation	21.19	24.68
Skew	0.13	0.51
Skew Standard Error	0.62	0.62
Kurtosis	-0.96	-0.80
Kurtosis Standard Error	1.19	1.19
Shapiro-Wilk	0.96	0.92
- p-value	0.81	0.28
Significance		
T-Test		1.61
- p-value		0.13

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years

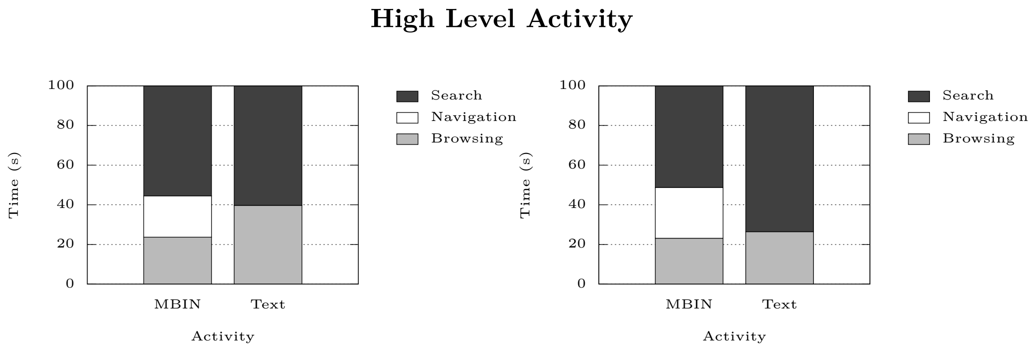


(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	7.00	7.00
Mean	36.81	34.70
Standard Deviation	14.95	24.07
Skew	-0.26	0.96
Skew Standard Error	0.79	0.79
Kurtosis	-0.73	-0.85
Kurtosis Standard Error	1.59	1.59
Shapiro-Wilk	0.94	0.77
- p-value	0.60	0.02
Significance		
T-Test		0.21
- p-value		0.84

(d) Statistics for the Total Time for question completion for both interfaces

Figure A.19: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

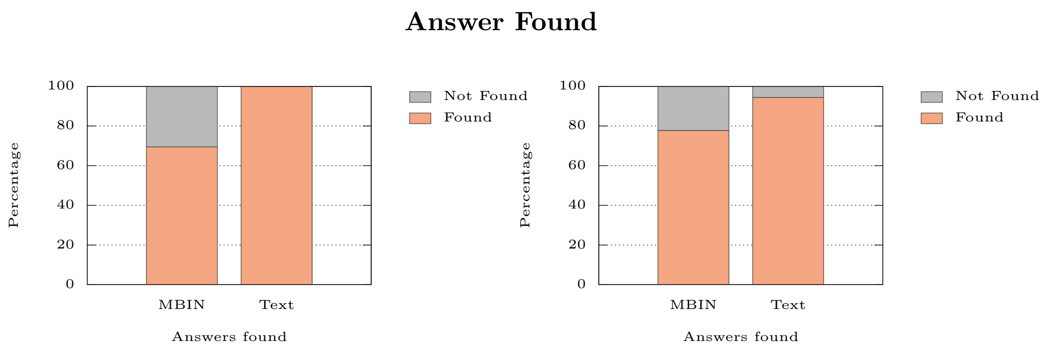


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.20: High level activity for below (a) and above (b) 2.5 years experience

(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.21: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

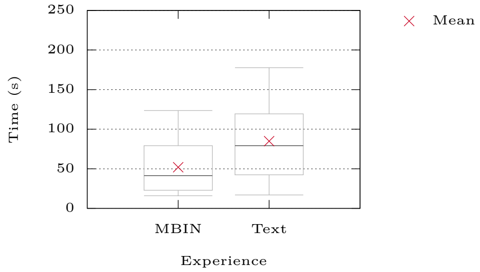


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.22: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.6 Task 6: Provide Technical Drawing; Instruct on Assembly Method; Provide Design Instruction

Less than 2.5 years

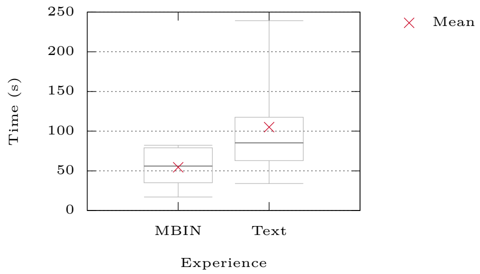


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	15.00	15.00
Mean	51.91	84.88
Standard Deviation	33.60	51.80
Skew	0.74	0.52
Skew Standard Error	0.58	0.58
Kurtosis	-0.81	-0.98
Kurtosis Standard Error	1.12	1.12
Shapiro-Wilk	0.87	0.91
- p-value	0.03	0.16
Significance		
T-Test		-2.39
- p-value		0.03

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years



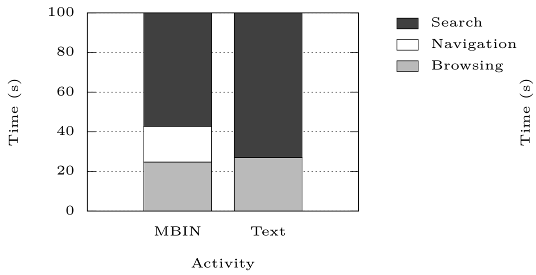
(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	8.00	8.00
Mean	54.48	105.05
Standard Deviation	23.72	63.54
Skew	-0.20	1.07
Skew Standard Error	0.75	0.75
Kurtosis	-1.50	-0.10
Kurtosis Standard Error	1.48	1.48
Shapiro-Wilk	0.90	0.86
- p-value	0.30	0.12
Significance		
T-Test		-2.73
- p-value		0.03

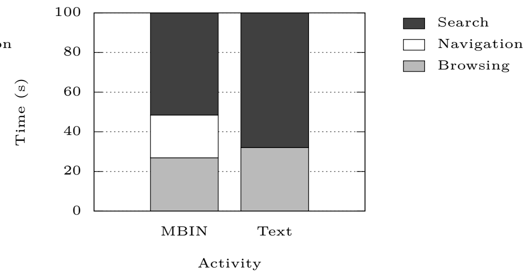
(d) Statistics for the Total Time for question completion for both interfaces

Figure A.23: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

High Level Activity



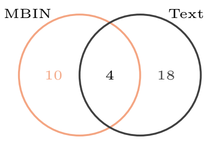
(a) Less than 2.5 years



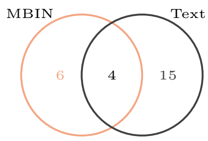
(b) Greater than 2.5 years

Figure A.24: High level activity for below (a) and above (b) 2.5 years experience

Reports Opened



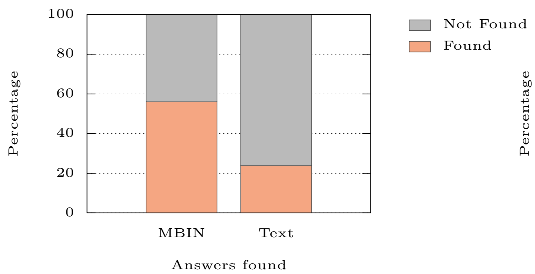
(a) Less than 2.5 years



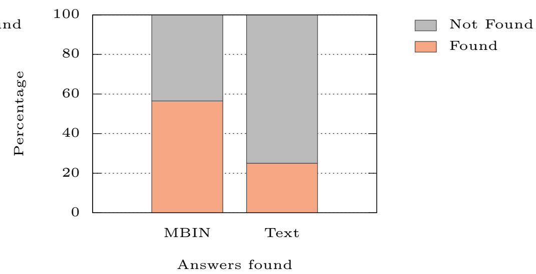
(b) Greater than 2.5 years

Figure A.25: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

Answer Found



(a) Less than 2.5 years

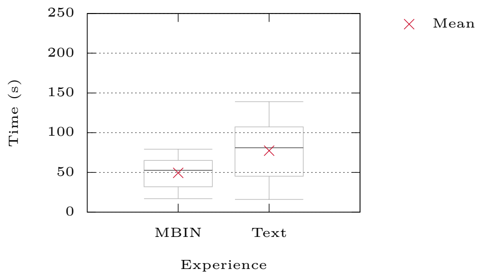


(b) Greater than 2.5 years

Figure A.26: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.7 Task 7: Provide Aerodynamics Analysis; Search for Similar Report

Less than 2.5 years

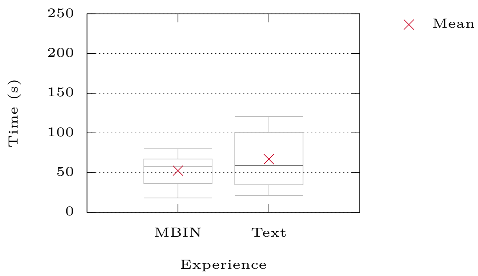


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	14.00	14.00
Mean	49.49	77.28
Standard Deviation	20.22	39.08
Skew	-0.24	-0.21
Skew Standard Error	0.60	0.60
Kurtosis	-1.26	-1.22
Kurtosis Standard Error	1.15	1.15
Shapiro-Wilk	0.93	0.93
- p-value	0.32	0.36
Significance		
T-Test		-2.33
- p-value		0.04

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years



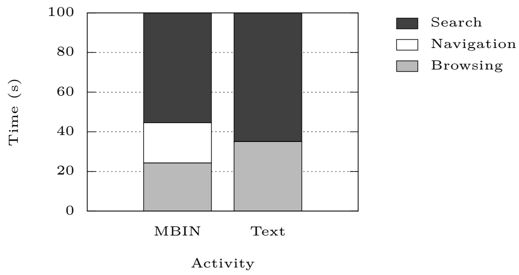
(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	8.00	8.00
Mean	52.53	66.84
Standard Deviation	19.64	35.42
Skew	-0.36	0.26
Skew Standard Error	0.75	0.75
Kurtosis	-1.15	-1.47
Kurtosis Standard Error	1.48	1.48
Shapiro-Wilk	0.94	0.91
- p-value	0.65	0.35
Significance		
T-Test		-0.93
- p-value		0.38

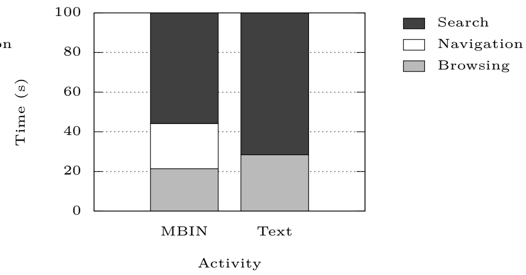
(d) Statistics for the Total Time for question completion for both interfaces

Figure A.27: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

High Level Activity



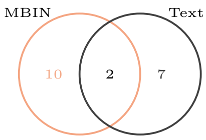
(a) Less than 2.5 years



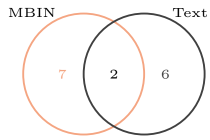
(b) Greater than 2.5 years

Figure A.28: High level activity for below (a) and above (b) 2.5 years experience

Reports Opened



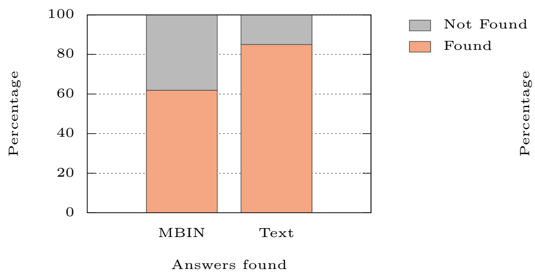
(a) Less than 2.5 years



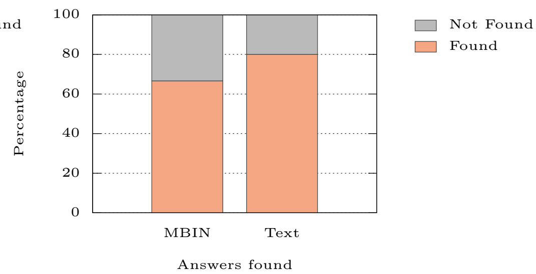
(b) Greater than 2.5 years

Figure A.29: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

Answer Found



(a) Less than 2.5 years

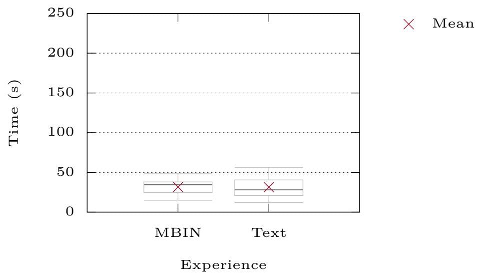


(b) Greater than 2.5 years

Figure A.30: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.8 Task 8: Identify Areas of the Case Worked on by an Individual

Less than 2.5 years

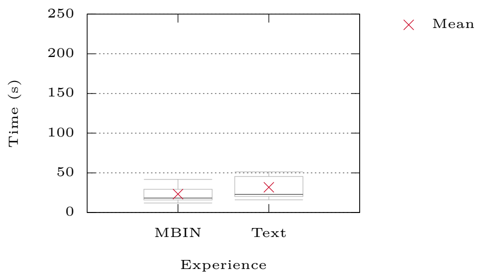


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	12.00	12.00
Mean	31.59	31.41
Standard Deviation	9.55	12.81
Skew	-0.27	0.40
Skew Standard Error	0.64	0.64
Kurtosis	-0.86	-0.88
Kurtosis Standard Error	1.23	1.23
Shapiro-Wilk	0.94	0.96
- p-value	0.48	0.81
Significance		
T-Test		0.05
- p-value		0.96

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years



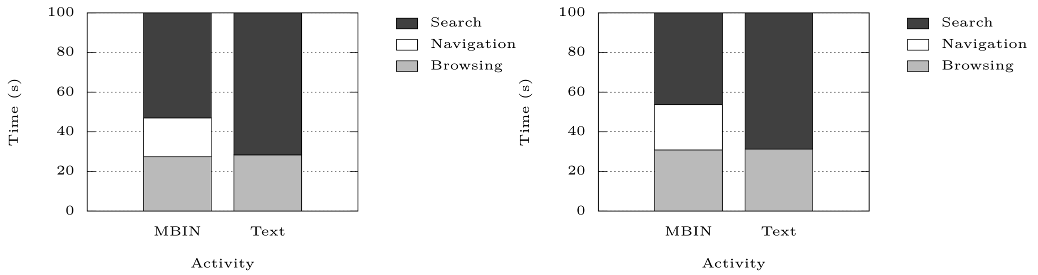
(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	7.00	7.00
Mean	23.22	31.70
Standard Deviation	9.75	13.86
Skew	0.69	0.32
Skew Standard Error	0.79	0.79
Kurtosis	-0.80	-1.74
Kurtosis Standard Error	1.59	1.59
Shapiro-Wilk	0.90	0.82
- p-value	0.32	0.07
Significance		
T-Test		-1.23
- p-value		0.26

(d) Statistics for the Total Time for question completion for both interfaces

Figure A.31: Total Time taken for below (a and b) and above (c and d) 2.5 years experience

High Level Activity



(a) Less than 2.5 years (b) Greater than 2.5 years

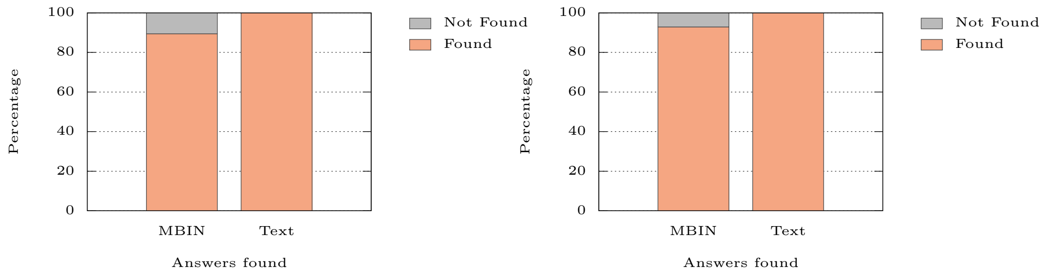
Figure A.32: High level activity for below (a) and above (b) 2.5 years experience

Reports Opened

(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.33: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience

Answer Found

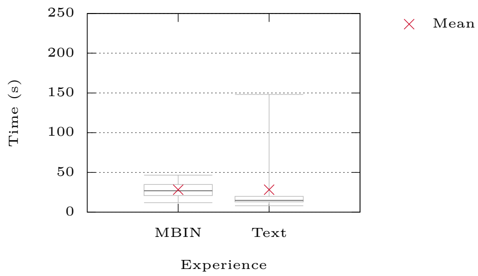


(a) Less than 2.5 years (b) Greater than 2.5 years

Figure A.34: Ability to answer question for below (a) and above (b) 2.5 years experience

A.4.9 Task 9: Identify New Knowledge from a Seemingly Unconnected Source; Report Manufacture Issue

Less than 2.5 years

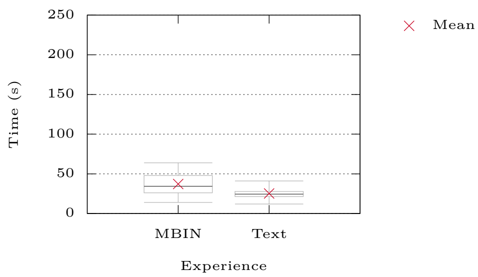


(a) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	16.00	16.00
Mean	50.02	27.63
Standard Deviation	22.41	12.98
Skew	0.07	0.74
Skew Standard Error	0.56	0.56
Kurtosis	-1.34	-0.49
Kurtosis Standard Error	1.09	1.09
Shapiro-Wilk	0.94	0.92
- p-value	0.30	0.18
Significance		
T-Test		3.05
- p-value		0.01

(b) Statistics for the Total Time for question completion for both interfaces

Greater than 2.5 years

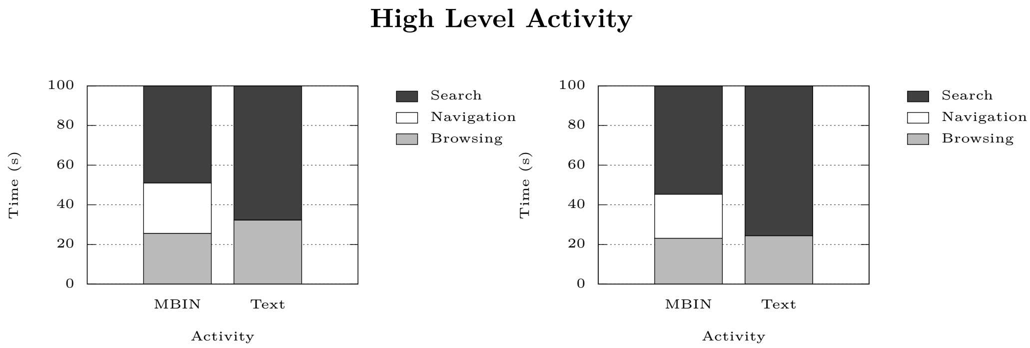


(c) Total Time for question completion for both interfaces

Measure	MBIN	Text
Sample Size	7.00	7.00
Mean	37.23	25.18
Standard Deviation	15.80	8.21
Skew	0.28	0.51
Skew Standard Error	0.79	0.79
Kurtosis	-1.04	-0.21
Kurtosis Standard Error	1.59	1.59
Shapiro-Wilk	0.98	0.95
- p-value	0.94	0.74
Significance		
T-Test		1.43
- p-value		0.20

(d) Statistics for the Total Time for question completion for both interfaces

Figure A.35: Total Time taken for below (a and b) and above (c and d) 2.5 years experience



(a) Less than 2.5 years

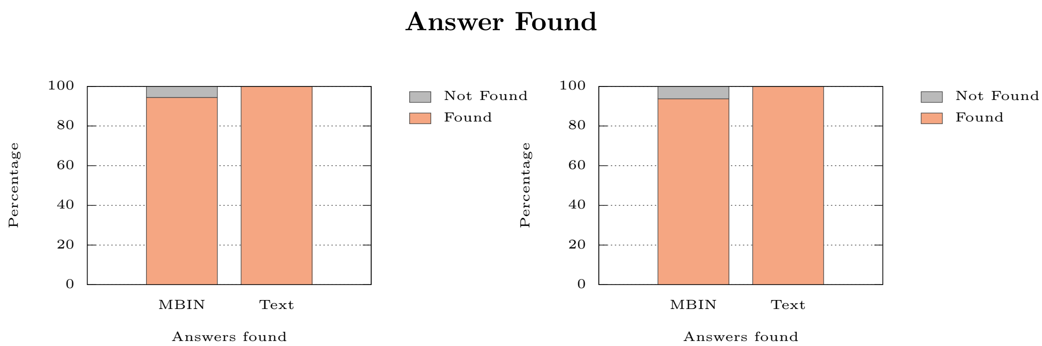
(b) Greater than 2.5 years

Figure A.36: High level activity for below (a) and above (b) 2.5 years experience

(a) Less than 2.5 years

(b) Greater than 2.5 years

Figure A.37: Venn diagram of the total number of reports opened in each interface for below (a) and above (b) 2.5 years experience



(a) Less than 2.5 years

(b) Greater than 2.5 years

Figure A.38: Ability to answer question for below (a) and above (b) 2.5 years experience